COMPARISON OF STATISTICAL AND NEURAL NETWORK METHODS FOR TIME-SERIES FOR FORECASTING

by
B Prasad

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLGOY, KANPUR

February 2000

COMPARISON OF STATISTICAL AND NEURAL NETWORK METHODS FOR TIME-SERIES FOR FORECASTING

A Thesis Submitted
in partial fulfilment of the Requirement
for the degree of

MASTER OF TECHNOLOGY

by

B Prasad

to the

DEPARTMENT OF ELECTRICAL ENGINEERING
INDIAN INSTITUTE OF TECHNOLGOY, KANPUR

February 2000

CERTIFICATE

It is to certify that work contained in the thesis entitled **Comparison of Statistical and Neural Network Methods for Time Series for Times Series Forecasting** ' by B Prasad has been carried out under my supervision and that has not been submitted else where for a degree

Dr Prem Kumar Kalra
Professor
Department of Electrical Engineering
Indian Institute of Technology Kanpur

ABSTRACT

One of the greatest challenges for human beings is to perceive the future so that we can get ourselves prepared for it The future of a process or a phenomenon depends on the past observations which are used to construct the time series forecasting model Traditionally statistical models or stochastic models were employed to model a time series The recent trend is applying Artificial Neural Network methods

In this present work a comparison of the performances of the statistical and Neural Network methods for time series forecasting is presented for some classical problems Box Jenkins approach is used for the statistical modelling The Neural Network models studied are Back Propagation through time and Time delayed Neural Networks

Dedicated

To

My parents

ACKNOWLEDGEMENT

I would like to express my gratitude to Dr P K Kalra my thesis supervisor for his inspiring guidance and sincere supervision not only during the thesis work but also during the complete academic programme I am also grateful to him personally for he has always helped me more than what I expected from him I would also like to thank him for the complete freedom afforded to me in my work as well as for the excellent facilities in the lab The completely open atmosphere for exchanging views in the lab not only with him but with all the others was what helped me in achieving what I could in this short span of time

My thanks to my labmates Murali Mohan Sinhaji Murali krishna Sudheer Tanmay Prasanth Madhav Krishna Maj Sidana for co operating in my work

I would like to thank all my friends of our hall shanki swaroop pav don chillar moorthy gudur kovvi sur gsrao chiru gopal professor who made my short stay at IIT Kanpur an enjoyable and memorable one Special thanks to my friend Masa shanki who helped me in my thesis documentation

Contents

chapter 1

INTRODUCTION

1 1 Introduction

It is human nature to know in advance what is likely to happen in the future Observing past outcomes of a phenomenon in order to anticipate its future behaviour represents the essence of forecasting or prediction If a complete mathematical model describing a studied phenomenon is known and not very complex and if the initial conditions are sufficiently defined forecasting becomes a trivial task But when an analytical model is unknown too complex then a typical alternative is to try to forecast by building a model that takes into account only previous outcomes of the phenomenon while ignoring any exterior influence

Forecasting is predicting the short term evolution of the system or a phenomenon Forecasting a system or a natural phenomenon is of utmost importance They help in planning and also in preventing a forecasted disaster Some of the major applications of forecasting include forecasting weather rainfall natural calamities sales and stock market population electric load demand

The outcomes of the phenomenon over time form a *time-series* A time-series can be defined as a function x of an independent variable t form a process for which a mathematical description is unknown Time series prediction problems are approached either from a stochastic perspective or more recently from a neural network perspective Each of these methods has their own advantages and disadvantages

Statistical Methods

The statistical models include models such as Auto regressive (AR) Moving Average (MA) or the combination of the two Auto Regressive Moving Average (ARMA) model These models have limited applicability as they commonly employ linear models The advantage of these models is that they are fast The commonly used statistical model in literature is that proposed by Box-Jenkins[1]

Neural Network Methods

Among the various potential applications of neural networks forecasting is considered to be a major application The Neural Network models are powerful with regard to the accuracy of prediction

The Back Propagation algorithm is the most popular method for the design of neural networks However it doesn t incorporate the dynamical behaviour which is a must for the forecasting problems Some Neural Network models which incorporate this dynamical property to it includes Time Delay Neural Networks Temporal back propagation Back propagation through time This work employs two Neural Network methods for forecasting a time series Time Delay Neural Networks[6] and Back propagation through time[2] A statistical method is also used to model the time series and its accuracy of prediction compared with the Neural Network methods

1 2 Problem Definition

This thesis compares the *forecasting performance* of a time series using two approaches the traditional Box Jenkins statistical method and the Neural Network methods to a time-series model The two Neural Network models discussed are back propagation through time and time delay neural networks A comparison of the two approaches is presented by taking some classical problems of time series modelling

1 3 Organisation of the thesis

A detailed discussion of the statistical approach of the Box Jenkins model is presented in chapter 2 wherein all the steps of model building are described It describes the different stages of model building using Box Jenkins method Chapter 3 has the concepts and the algorithms of back propagation through time and Time delay Neural Networks Results of all the above methods for some classical examples like the sunspots series save rate data are presented in Chapter 4 In the end conclusions and scope for future work are briefed in chapter 5

Chapter 2

BOX-JENKINS METHOD

An important part of statistics deals with the analysis of data that are collected sequentially over time or *time series data* There are two objectives of analyzing time series data One to model the historical data and two to forecast or predict future values of the series The technique that is discussed here is that of a single series which means that the model and forecasts are based only on past values of the variable being forecast These models are termed as UBJ (Univariate Box Jenkins) and are also referred to as UBJ ARIMA ARIMA stands for Auto Regressive Integrated Moving Average

All statistical forecasting methods are extrapolative in nature they involve the projection of past patterns or relationships into the future In the case of UBJ ARIMA forecasting we extrapolate past patterns within a single data series into the future In other words the model is an algebraic statement telling how one thing is statistically related to one or more other things

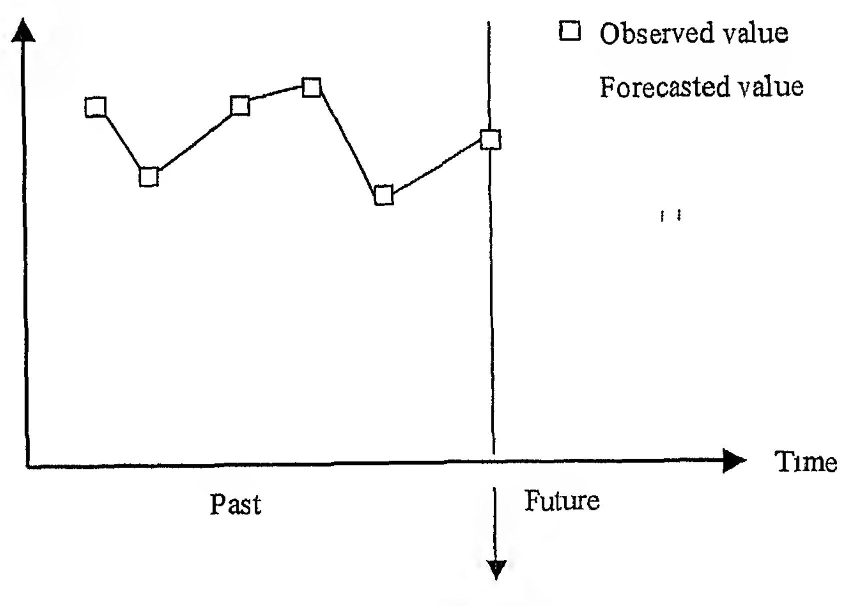


Fig 2 1 The idea of forecasting

2.1 Single-series (univariate) analysis

Time series analysis is used to explain the behavior of time series data using only past observations on the variable in question. The various observations in a time sequenced data (z_{t-1}, z_t, z_{t+1}) are assumed to be *statistically dependent*. The concept of correlation is used to measure the relationships between observations within the series. Figure shows the idea of UBJ forecasting.

2.2 When may UBJ model be used?

Short term forecasting

UBJ ARIMA models are especially suited to short term forecasting because most ARIMA models place heavy emphasis on the recent past rather than the distant past. The long tem forecasts are less reliable because the observations are not available and they themselves are predicted and hence less reliable.

Data types

The UBJ method applies to either *discrete data* or *continuous data*. The data must be equally spaced at discrete time intervals.

Sample size

Building an ARIMA model requires an adequate sample size, atleast 50 observations.

Stationary series

The UBJ ARIMA method applies only to *stationary* series which has a mean, variance and autocorrelation function that are essentially constant through time.

2.3 Statistical terms used

2.3.1 Differencing

Non stationary series, series for which the mean changes over time, can be transformed into stationary ones by employing differencing.

$$w_t = z_t - z_{t-1} \qquad t = 2, 3, \ldots, n \qquad (2.1)$$

The series w_t is called the *first differences* of z_t. If the first differences of z_t do not have a constant mean, then w_t is redefined as the first differences

$$w_t = (z_t - z_{t-1}) - (z_{t-1} - z_{t-2}) \qquad t = 3, 4, \ldots, n \qquad (2.2)$$

Series w_t is now called the *second differences* of z_t.

2 3 2 Deviations fi om the mean

To focus on the stochastic behavior of the stationary series the data is expressed in deviations from the mean i e we define a series $\bar{z}_t$ The two series z_t and $\bar{z}_t$ have all the same statistical properties except for their means

2 3 3 Estimated autocoi relation functions

Autocorrelation coefficient means calculating the correlation coefficient between sets of ordered pairs $(\tilde{z}_t\ \tilde{z}_{t\ 1})$ These autocorrelation coefficients when plotted graphically is known as autocorrelation function (acf) It measures the direction and strength of the statistical relationship between ordered pairs of observations on two random variables and can take values between -1 and $+1$ A value of 1 means perfect negative correlation, a value of $+1$ means perfect positive correlation and a value of 0 denotes uncorrelated

The standard formula for calculating autocorrelation coefficients is

$$r_k = \frac{\sum_{t=1}^{k}(z_k - \bar{z})(z_{t+k} - \bar{z})}{\sum_{t=1}(z_t - \bar{z})^2} \tag{2 3}$$

2 3 4 Estimated partial autocoi relation funtions

The estimated partial autocorrelation function(pacf) is broadly similar to an estimated acf It is used as a guide along with the estimated acf in choosing one or more ARIMA models that might fit the available data The idea of partial autocorrelation analysis is that we want to measure how $\tilde{z}_t$ and $\tilde{z}_{t+k}$ are related but with the effects of the intervening *z s accounted for* The partial autocorrelation coefficients are found by applying regression techniques

A computationally easier way to estimate ϕ_{kk} coefficients is

$$\phi_{kk} = r_1 \tag{2 4}$$

$$\phi_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \phi_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} r_j} \qquad (k = 2, 3, \ldots) \qquad (2.5)$$

where

$$\phi_{kj} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j} \qquad (k = 3, 4, \ldots;\ j = 1, 2, \ldots, k-1)$$

2.4 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) MODELS

There are three common ARIMA processes in literature. They are

1. Autoregressive (AR) Model
2. Moving Average (MA) Model
3. Autoregressive Moving Average (ARMA) Model

The ARIMA model can be represented in terms of (p, d, q) notation, where p indicates the autoregressive order, q the moving average order and d the degree of differencing necessary to achieve stationary.

Stationarity requirement

The UBJ ARIMA models are applicable to only those series which are stationary. If the time series is non stationary then it is made a stationary series by differencing.

Invertibility requirement

This ensures that the smaller weights are assigned to observations that are further in past.

2.4.1 Autoregressive Models

In *autoregressive processes* of order p, the current observation z_t is expressed as the sum of three components: a linear combination of the p immediate past observations, a constant C and a random error component for the current period. The basic model is represented by ARIMA(1,0,0) or mathematically as

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \ldots + \phi_p z_{t-p} + C + a_t \qquad (2.6)$$

where $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive parameters. The term autoregressive is used since z_t is regressed on observations from the same series.

The model in Eq (2.6) can be rewritten as

$$(1 \quad \phi_1 B \quad \phi_2 B^2 \qquad \phi_p B^p)\, z_t = C + a_t \tag{2 7}$$

$$(1 - \phi_1 B \quad \phi_2 B^2 \qquad \phi_p B^p) = \phi(B) \tag{2 8}$$

Table 2 1 Summary of stationarity conditions for AR coefficients

Model Type	Stationarity Conditions
ARMA(0 q)	Always stationary
AR(1) or ARMA(1 q)	$\lvert \phi_1 \rvert < 1$
AR(2) or ARMA(2 q)	$\lvert \phi_2 \rvert < 1$
	$\phi_2 + \phi_1 < 1$
	$\phi_2 \quad \phi_1 < 1$

2 4 2 Moving Average Models

In moving average processes of order q the current observations z_t is the sum of a current and weighted lagged random error terms for the last q periods together with a constant C The ARIMA (0 0 q) or MA (q) process is determined by

$$z_t = C \quad a_t \quad \theta_1 a_{t\,1} \quad \theta_2 a_{t\,2} \qquad \theta_q a_{t\,q} \tag{2 9}$$

or

$$z_t = C + (1 \quad \theta_1 B \quad \theta_2 B^2 \qquad \theta_q B^q)\, a_t \tag{2 10}$$

where

$$(1 \quad \theta_1 B \quad \theta_2 B^2 \qquad \theta_q B^q) = \theta(B) \tag{2 11}$$

Table 2 2 Summary of invertibility conditions for MA coefficients

Model Type	Invertibility Conditions
ARMA(p 0)	Always invertible
MA(1) or ARMA(p 1)	$\theta_1 < 1$
MA(2) or ARMA(p 2)	$\theta_2 < 1$
	$\theta_2 + \theta_1 < 1$
	$\theta_2 \quad \theta_1 < 1$

2 4 3 Autoregressive Moving Average (ARMA) models

A time series can have both autoregressive and moving average terms and so has both their features For example consider the ARMA(1 1) model given by

$$z_t = C + \phi_1 z_{t-1} + a_t - \theta_1 a_{t-1} \qquad (2\ 12)$$

or

$$(1 - \phi_1 B) z_t = C + (1 - \theta_1 B) a_t \qquad (2\ 13)$$

The mixed autoregressive moving average process of order $(p\ q)$ denoted by ARMA $(p\ q)$ is represented by the equation 2 13

2 5 The Box Jenkins modeling procedure

Box and Jenkins propose a practical three stage procedure for finding a good model The broad outline of modeling strategy is shown in the flow chart below The three stages are

1 Identification

2 Estimation

3 Diagnostic checking

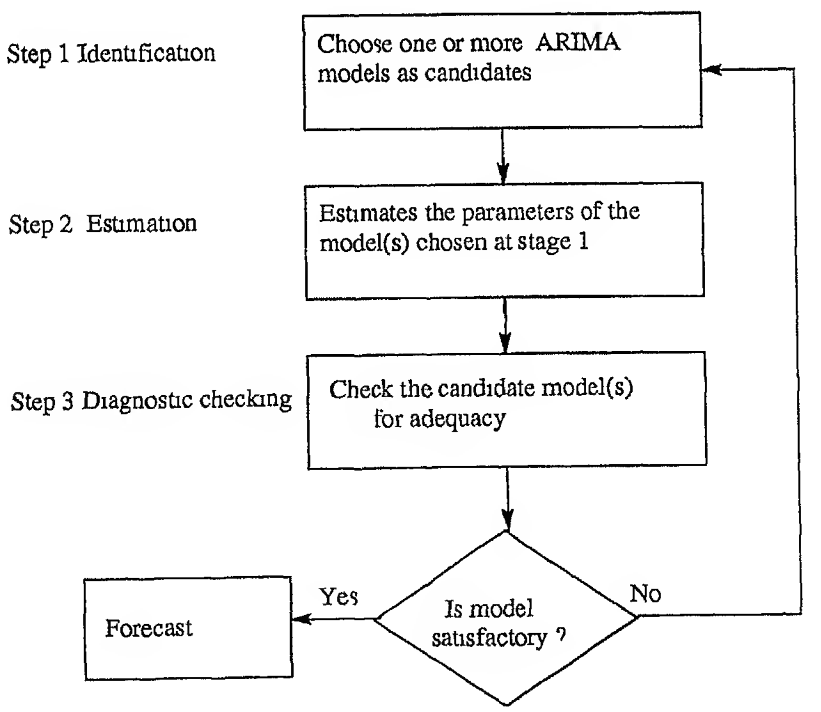


Fig 2 2 Stages in Box Jenkins iterative approach to model building

2 5 1 What is a good model?

There is a difference between a model and a process A process is the true but unknown mechanism that has generated a realisation, while a model is only an imitation or representation of the process Because the process is unknown we never know if the selected model is the correct one The following are the characteristics of a good model

1 A good model is *parsimonious*
2 A good AR model is *stationary*
3 A good MA model is *invertible*
4 A good model has *high quality estimated coefficients* at the estimation stage
5 A good model has *statistically independent residuals*
6 A good model *fits the available data sufficiently well* at the estimation stage
7 Above all a good model has *sufficiently small forecast errors*

2 6 Model Building

2 6 1 Identification

At this stage it is necessary to identify the values of (p d q) This identification is solely based on the examination of the data We use two graphical devices the estimated acf and the estimated pacf to identify the underlying model The basic idea in this identification is that every ARIMA model has a *theoretical* acf and pacf associated with it At the identification stage we compare the *estimated* acf and pacf calculated from the available data with various *theoretical* acf's and pacf's We then tentatively choose the model whose theoretical acf and pacf most closely resemble the estimated acf and pacf of the data series The statistical test such as the *t tests* or the *chi squared test* on the acf and the pacf is used to identify the tentative model Whichever model we choose at the identification stage we consider it only tentatively and it is only a candidate for the final model

Theoritical acf and pacf for some common processes

The major characteristics of theoretical acf's and pacf's for stationary AR, MA and mixed (ARMA) processes

Table 2 3 Primary distinguishing characteristics of theoretical acf's and pacf's for stationary processes

Process	acf	pacf
AR	Tails off towards zero (exponential decay or damped sine wave)	Cuts off to zero (after lag p)
MA	Cuts off to zero (after lag q) (exponential decay or damped sine wave)	Tails off toward zero
ARMA	Tails off toward zero	Tails off toward zero

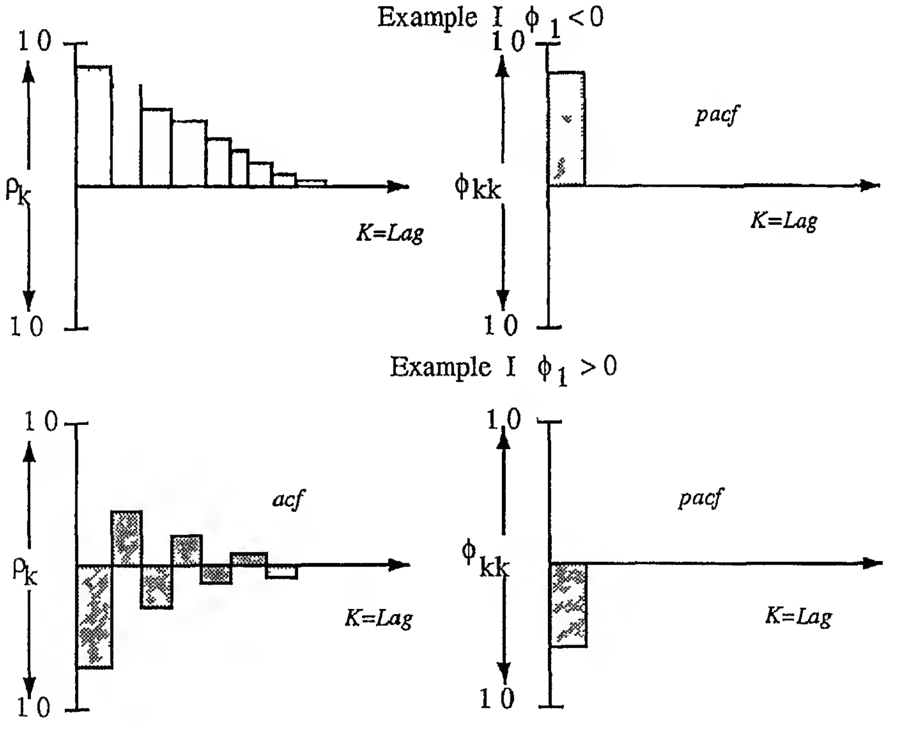


Fig 2 3 Examples of theoretical acf and pacf for two stationary AR(1) processes

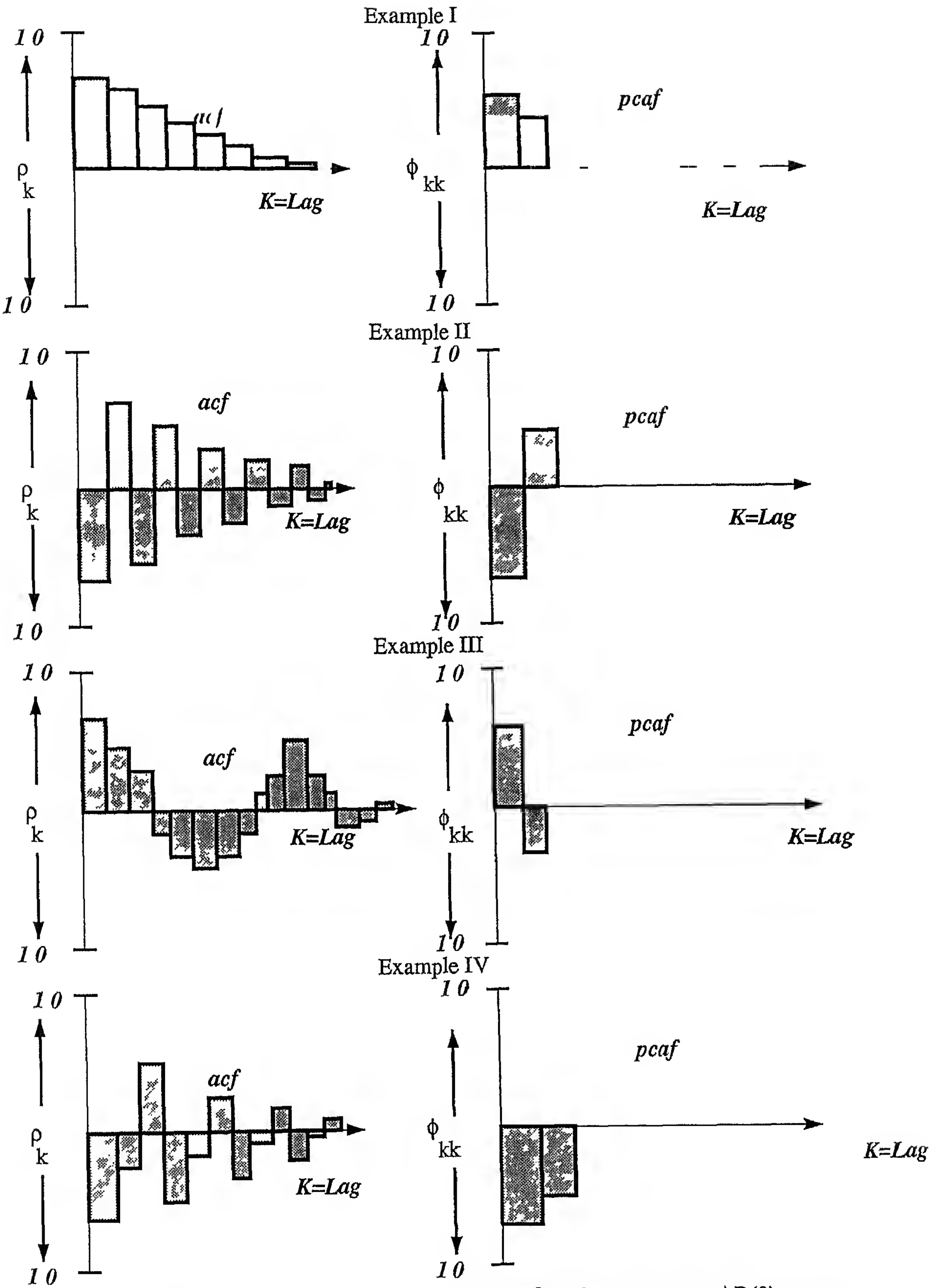


Fig 2 4 Examples of theoritical acf and pacf for four stationary AR(2) processes

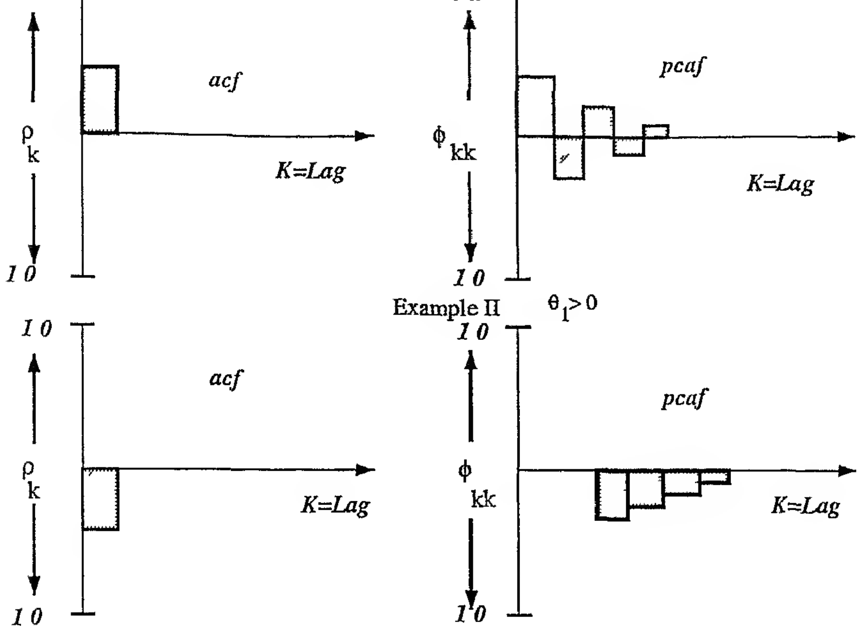


Fig 2 5 Examples of theoretical acf and pacf for two MA(1) processes

Table 2 4 Detailed characteristics of some common stationary processes

Process	acf	pacf
AR (1)	Exponential decay (a) on the positive side if $\phi_1 > 0$ (b) alternating in sign on the negative	Spike at lag 1 then cuts off to zero (a) spike is positive if $\phi_1 < 0$ side if $\phi_1 < 0$
AR(2)	A mixture of exponential decays or a damped sine wave The exact pattern depends on the signs and sizes of ϕ_1 and ϕ_2	Spikes at lags 1 and 2 then cuts off to zero
MA(1)	Spike at lag 1 then cuts off to zero (a) spike is positive if $\theta_1 < 0$ (b) spike is negative if if $\theta_1 > 0$	Damps out exponentially (a) alternating in sign, starting on the positive side $\theta_1 < 0$ (b) on the negative side if $\theta_1>0$
MA(2)	Spikes at lags 1 and 2 then cuts off zero	A mixture of exponential decay to or a damped sine wave The exact pattern depends on the signs and sizes of θ_1 and θ_2
ARMA(1 1)	Exponential decay from lag 1 (a)sign of ρ_1=sign of (ϕ_1 θ_1) (b)all one sign if $\phi_1 > 0$ (c)alternating in sign if $\phi_1 < 0$	Exponential decay from lag 1 (a) $\phi_{11} = \rho_1$ (b)all one sign if θ_1 > 0 (c) alternating in sign if $\theta_1 < 0$

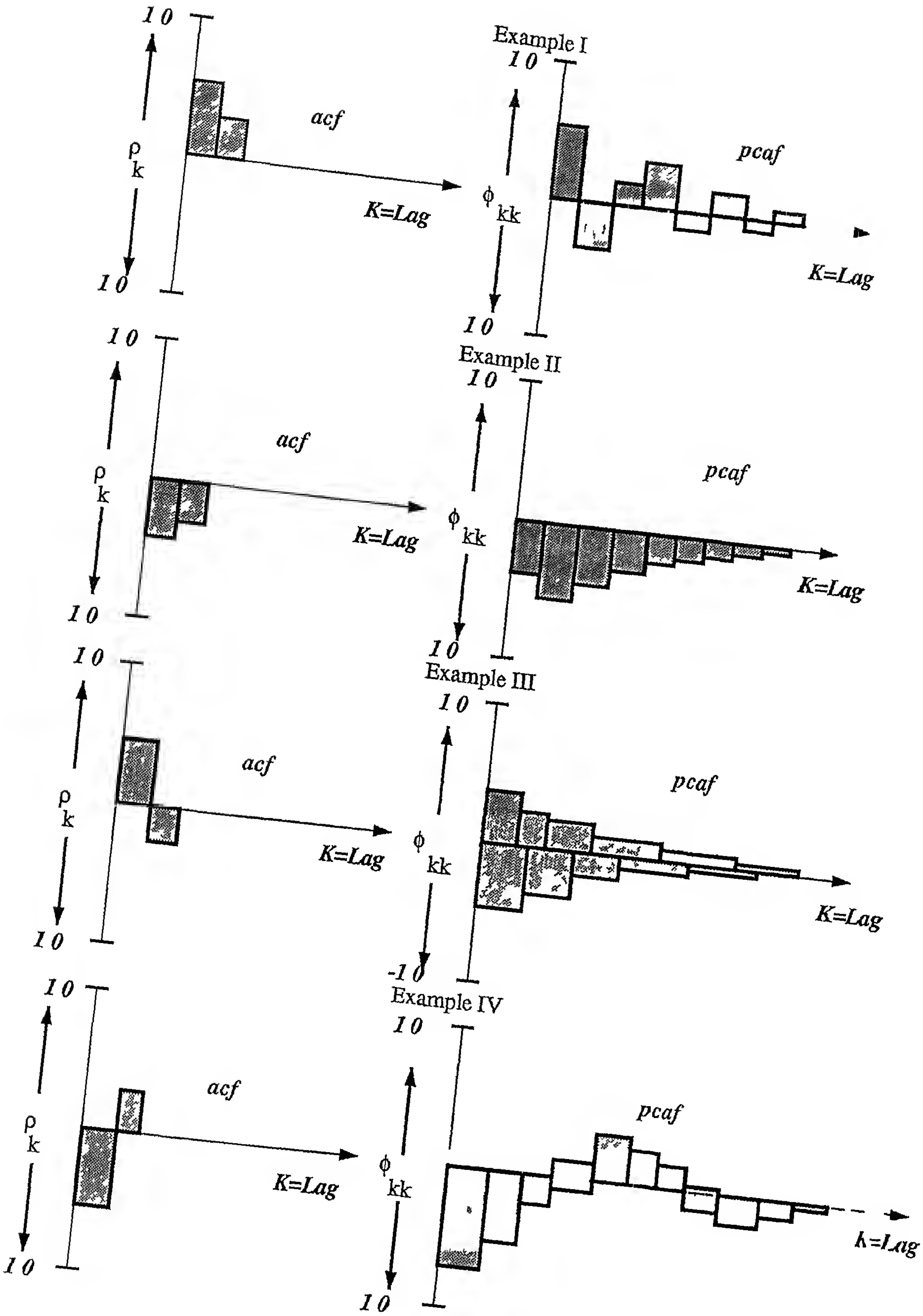


Fig 2 6 Examples of theoritical acf and pacf for four MA(2) processes

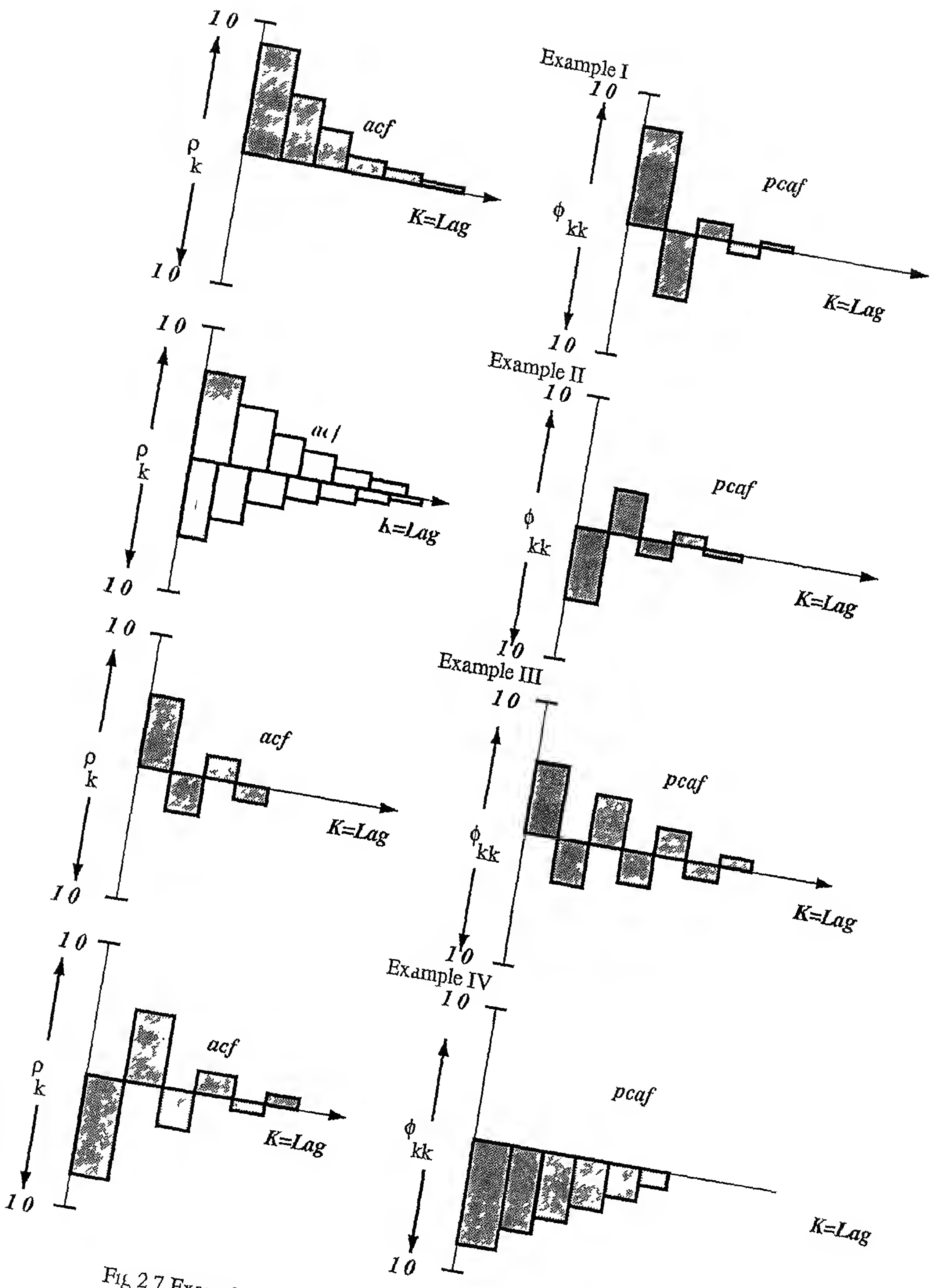


Fig 2 7 Examples of theoritical acf and pacf for two ARMA(1 1) processes

2 6 2 Estimation

Once a model is tentatively identified the parameters can be estimated by maximizing the corresponding likelihood function, assuming that the white noise tem is normally distributed At the estimation stage we get the precise estimates of a few parameters as we fit out tentative model to the data

Box and Jenkins favor choosing coefficient estimates at the estimation stage according to the maximum likelihood (ML) criterion But finding exact ML estimates can be computationally burdensome so Box and Jenkins suggest the use of least squares (LS) estimates If the random shocks are Normally distributed LS estimates are computationally easier to find and provide exactly or very nearly ML estimates LS estimates are those which give the smallest sum of squared residuals (SSR = a_t2)

A residual (a_t) is an estimate of a random shock (a_t) It is defined as the difference between an observed value (z_t) and a calculated value ($\overline{z}_t$) In practice the calculated values are found inserting estimates of the mean and the AR and MA coefficients into the ARIMA model being estimated with the current random shock assigned its expected value of zero and applying these estimates to the available data.

Linear least squares (LLS) may be used to estimate only pure AR models without multiplicative seasonal terms All other models require a nonlinear least squares (NLS) method

The most commonly used NLS method is the combination of two NLS procedures Guass Newton linearization and the gradient method This combination is sometimes called as Marquardt s compromise (the algorithm is shown in Fig 2 8) Given some initial estimates this algorithm chooses a series of optimal coefficients corrections This method converges quickly to LS values in most cases The estimated results may be used to check a model for stationarity and invertibility

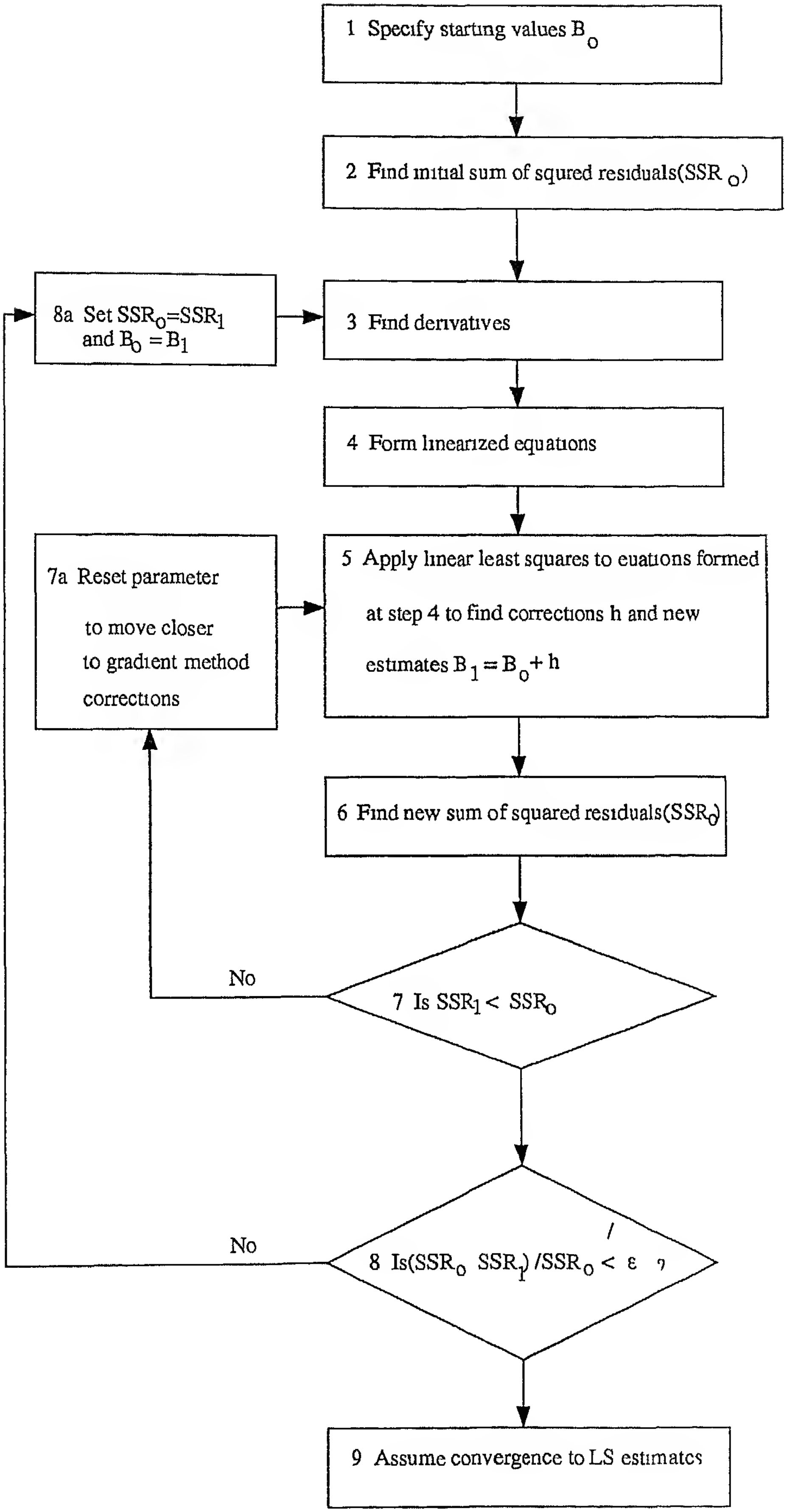


Fig 2 8 Flow chart for Marquardt s compromise for model estimation

2 6 2 Diagnostic Checking

Once we have obtained the precise estimates of the coefficients in an ARIMA model the third stage in the UBJ procedure diagnostic checking is done At this stage we decide if the estimated model is statistically adequate Diagnostic checking is related to the identification stage in two important ways First when the diagnostic checking shows that the model is inadequate then we must return to the identification stage to tentatively select one or more other models Second diagnostic checking also provides clues about how an inadequate model might be reformulated

The most important test of the statistical adequacy of an ARIMA model involves the assumption that the random shocks are independent If the random shocks are dependent or serially correlated it means that there is an auto correlation pattern in z_t that has not been accounted for by the AR and MA terms in that model

The basic analytical tool at the diagnostic checking stage is the residual acf The calculation of residual acf is the same as that of the estimated acf with the only difference that the residuals are taken instead of the actual realisation After calculating the residual autocorrelations its standard error and t tests are performed

Criterion employed

If the absolute value of a residual acf's *t value* is less than (roughly) 1 25 at lags 1 2 and 3 and less than about 1 6 at larger lags we can conclude that the random shocks at that lag are independent This is just an approximate guess on the adequacy of the model

2 6 4 Forecasting

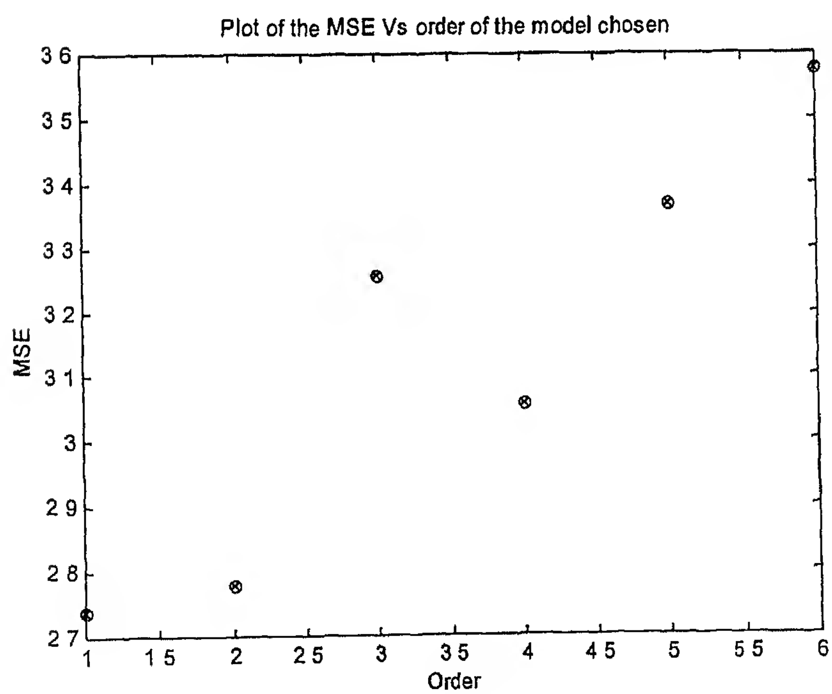
The ultimate application of UBJ ARIMA modeling is to forecast future values of a time series After estimating the parameters and doing the diagnostic checking for the chosen model forecasting is done The observations are available up to time t the forecast origin Thereafter they must be replaced by their forecasted values Past a_t values are also replaced by their corresponding estimates or the estimated residuals $\hat{a}_t$

2 7 Simulation Results

Example 1 Change in Business Inventories

The testing MSE obtained for different orders chosen is tabulated below This variation of testing MSE with respect to the order chosen is plotted graphically

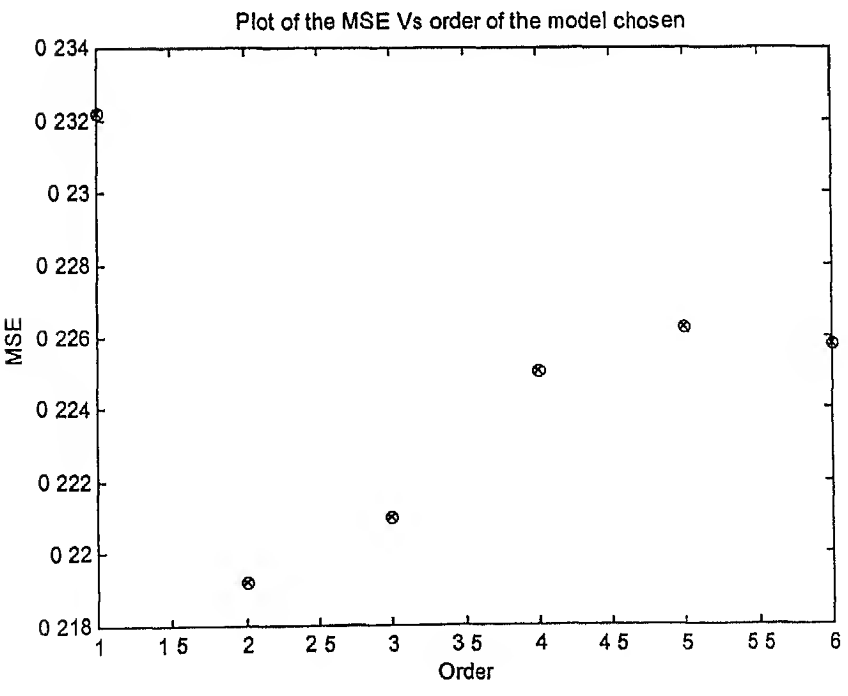
Order	Testing M S E
1	2 7401
2	2 7772
3	3 2569
4	3 0562
5	3 3678
6	3 5750



Example 2 Save Rate data

The testing MSE obtained for different orders chosen is tabulated below This variation of testing MSE with respect to the order chosen is plotted graphically

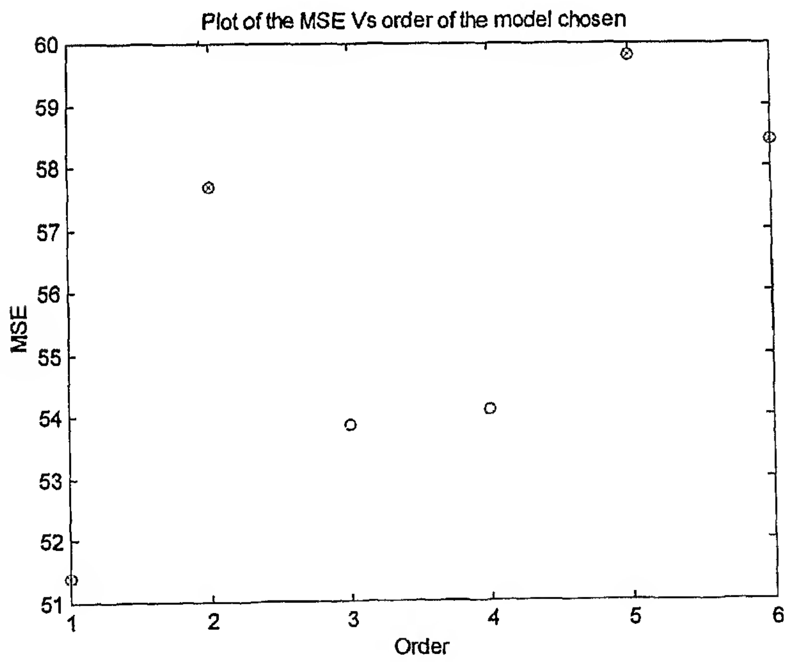
Order	M S E
1	0 23221
2	0 21924
3	0 22098
4	0 22502
5	0 22627
6	0 22579



Example 3 Sunspots data

The testıng MSE obtaıned for dıfferent orders chosen ıs tabulated below Thıs varıatıon of testıng MSE wıth respect to the order chosen ıs plotted graphıcally

Order	M S E
1	51 4028
2	57 6942
3	53 8622
4	54 0858
5	59 8086
6	58 4127



Chapter 3

NEURAL NETWORK METHODS

3 1 Introduction

One of the important and widely accepted applications of Neural Network is in the challenging field of *forecasting a trend* The highly complex and chaotic time series can be model to a good amount of accuracy by some Neural Networks

Static neural networks like the simple back propagation are trained to produce a spatial output pattern in response to a particular spatial input pattern However in many engineering scientific and economic applications the need arises to model dynamical processes where a time sequence is required in response to certain temporal input signal(s) The resulting model is referred to as a temporal association network Temporal associations must have a recurrent (as opposed to a static) architecture so as to handle the time dependent nature of associations Thus it would be very to extend the multilayer feedforward network and its associated training algorithms (like backprop) into the temporal domain In general it requires a recurrent architecture (nets with feedback connections) and proper associated learning algorithms Two such temporal networks discussed are Back Propagation Through Time (BPTT) and Time Delay Neural Networks (TDNN)

3 2 Back-Propagation Through Time

The back propagation through time (BPTT) algorithm for training a recurrent network is an extension of the standard back propagation algorithm

It is based on the fact that for every recurrent network there exits a feed forward network with identical behavior over a finite period of time This behavior of a recurrent network can be achieved unfolding the temporal operation of the network into a multilayer feed forward network The topology of such a network grows by one layer at every time step

A comparative representation of a simple recurrent network and a back propagation through time network is shown in the figure 3 1 The constraint for the above network in fig (3 1) is that the weights at each level of the feed forward network should be the same The appropriate method for maintaining this constraint is to keep track of the changes dictated for each weight at each level and then change each of the weights according to the sum of these individually prescribed changes

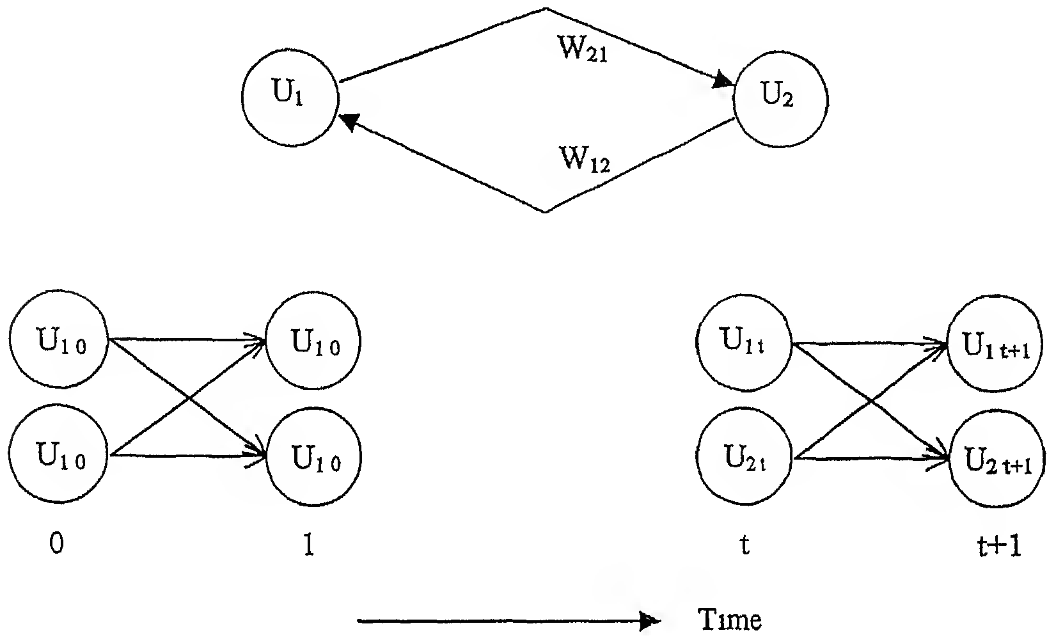


Fig 3 1 Comparison of the recurrent network and a feedforward network with identical behavior

The general rule for determining the change prescribed for a weight in the system for a particular time is that of the simple back propagation algorithm given below

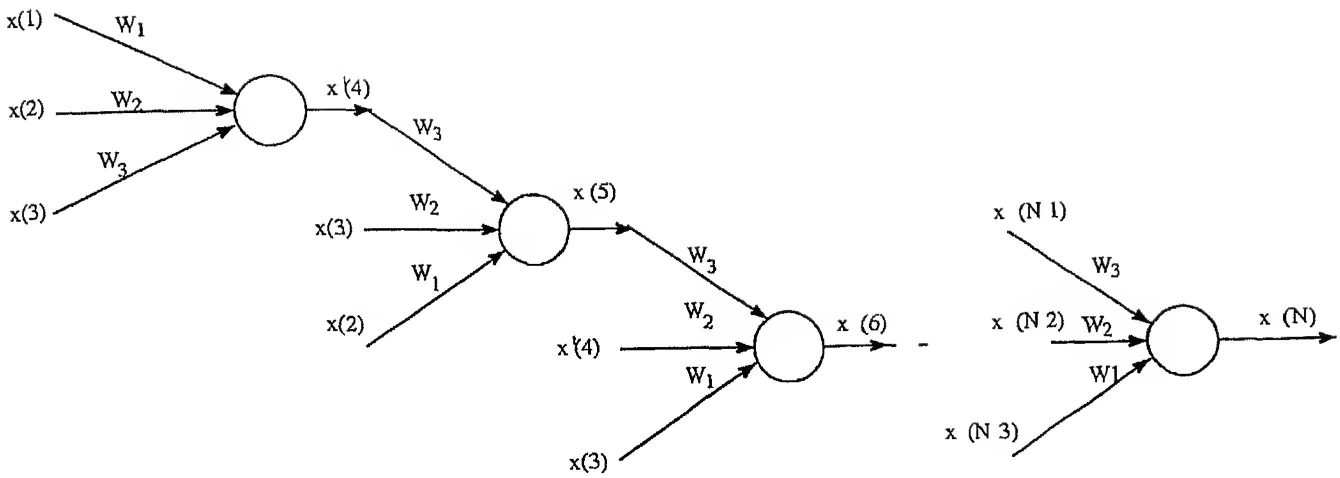


Fig 3 2 Back Propagation Through Time network

3 2 1 Simple Back Propagation method for weight updation

The weight updation used in the simple back propagation algorithm is as given below The mathematical equations governing the propagation of the error are as follows The schematic diagram of such a network is shown in Fig 3 3

The error function or cost function that is to be minimized is

$$E = 0\ 5 \sum_{k=1}^{K} (d_k - o_k)^2 \qquad (3\ 1)$$

The weights of the neuron whose desired outputs are known are updated as

$$W_{kj}^{n} = W_{kj} + \eta \delta_{k} y_j \qquad (3\ 2)$$

The weights of the inner layer neurons whose outputs are not known are updated as

$$V_{ji} = V_j + \eta \delta_{yj} z \qquad (3\ 3)$$

where

$$\delta_{ok} = (\delta_k - o_k) f\ (net_k) \qquad (3\ 4)$$

$$\delta_{yj} = f_j(net_j)\sum_{k=1}^{K}\delta_k W_{kj} \qquad (3\ 5)$$

The derivative of the activation function is given by

For *Unipolar activation functions*

$$f\ (net) = o_k(1-o_k) \qquad (3\ 6)$$

For Bipolar activation function

$$f\ (net) = 0\ 5(1-o_k)^2 \qquad (3\ 7)$$

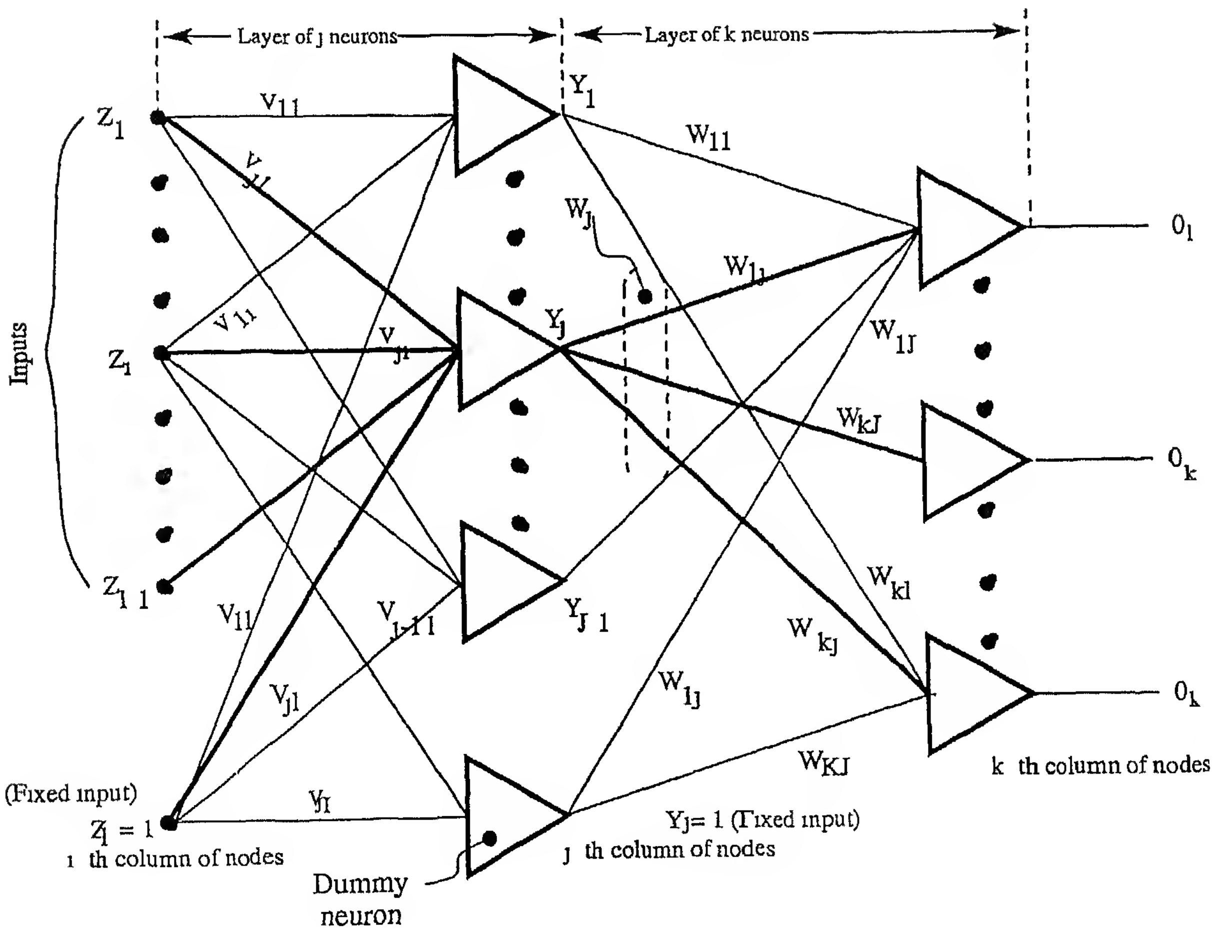


Fig 3 3 Feedforward Neural Network

3 2 2 Algorithm for BPTT

Step 1 choose the number of inputs or the past history that is to be submitted to the network

Step 2 An input is presented to the system with some initialised weights with the constraint that all the corresponding weights for the network remains the same

Step 3 The errors at all the neurons for which the desired output is known is calculated and added

Step 4 Weights are computed for all the units and the sum of all the weight changes dictated for a particular weight is saved

Step 5 The weights are changed by the amount of the sum of changes

Step 6 Go to step 1 until all the iterations are completed

3 3 Time Delay Neural Networks

Simple recurrent networks have been proven to be inadequate for several prediction tasks possibly because simple gradient procedures don t perform well in complex prediction tasks characterized by the existence of many local optima Now we can see one neural network model that have been successfully used for solving some problems in which values for a variable need to be predicted from past values for that variable

The generic network model shown in Fig 3 4 consists of preliminary preprocessing component that transforms an external input vector $x(t)$ into $\bar{x}(t)$

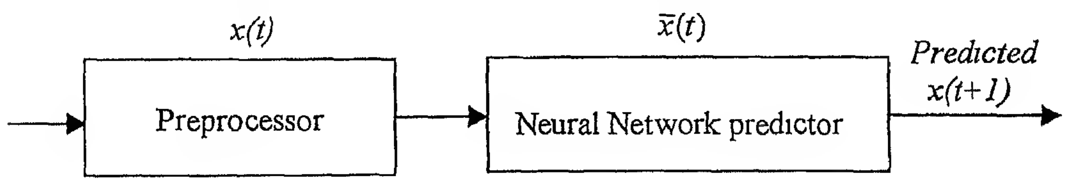


Fig 3 4 Generic neural network model for prediction

A preprocessed vector $\bar{x}(t)$ is supplied to a feed forward network The feed forward network is trained to compute the desired output values for a specific input $x(t)$ we have to note that $\bar{x}$ and x may be of different dimensionality The preprocessing

component may implement a short term memory and the feed forward network is the predictor component following the terminology of a mozer

Let us take a prediction task when *x(t)* was to predicted from *x(t 1) x(t 2)* In this simple case x at time t consists of a single input *x(t)* and $\bar{x}$ at time t consists of the vector (*x(t-1) x(t 2)*) supplied as input to feed forward network For this example preprocessing consists of merely of storing past values of the variable and supplying them to the network along with the latest value Such a model is sometimes called Tapped Delay line Neural Network consisting of a sequence of delay units or buffers and with the values of variable at recent instants being supplied to the feed forward predictor component

The architecture of this model is above in Fig 3 5 Algorithms such as back propagation and its variants are used to train the weights in the network and is given in 3 2 1

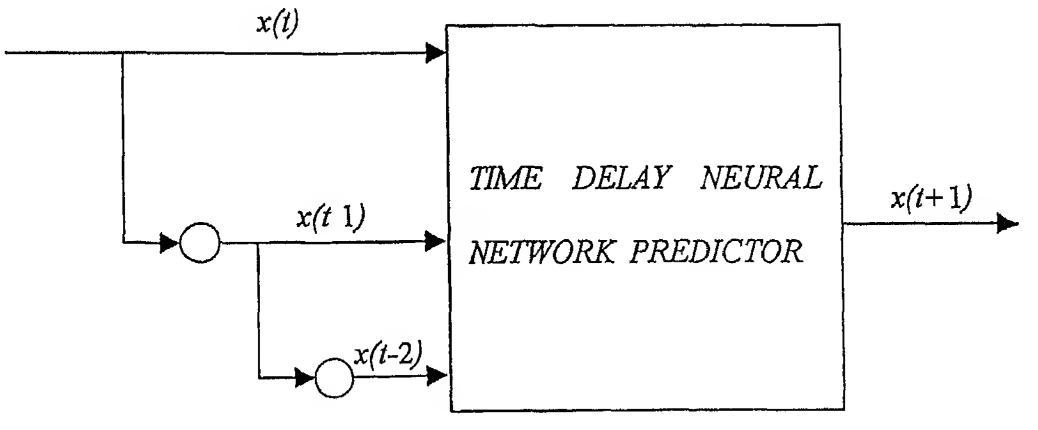


Fig 3 5 Time Delay Neural Network (TDNN) with two delay elements

Several questions need to be answered before attempting to forecast with an MLFF How many samples are required? How many input data points should be used that is what window size is best? What prediction horizon should be chosen (how far into the future to predict) how should the test and training data be divided what network configuration should be used (number of hidden layers number of nodes per layer type of activation functions type of output nodes) and so on Perhaps the most difficult question of all to answei is what to predict For example if the series is the Standard and

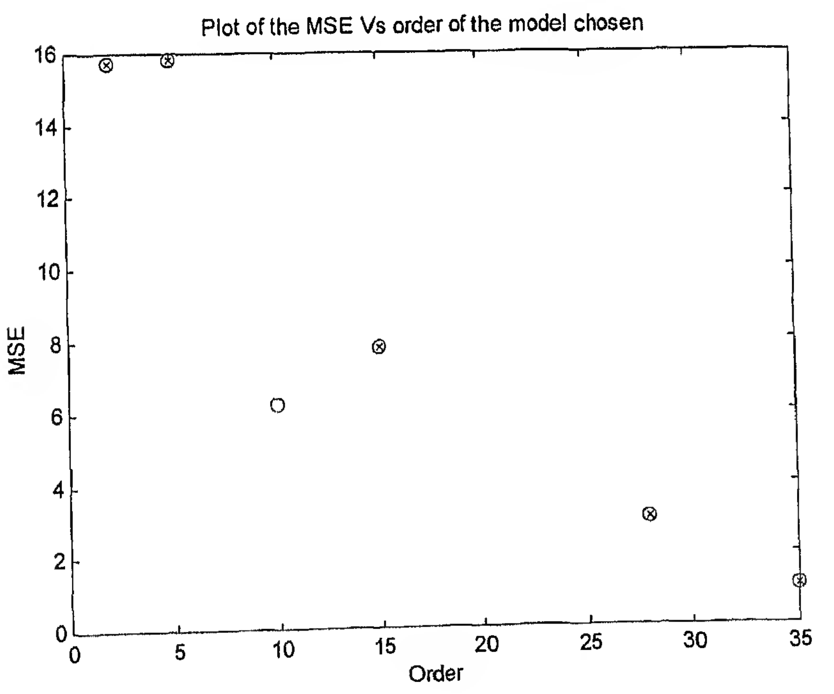
Poor (S&P) 500 stock index should the index value be forecast or should the direction of the series be forecast? The questions are not easily answered Knowledge of the dynamics driving the system to be forecast can provide some suggestions on good network architecture and network parameters Therefore if the user has tools to carry out statistical and chaotic data analyses it may be possible to characterize the system and provide some insight into the best choice of network parameters In any case one should set realistic objectives on what is achievable and then experiment with different architectures and training data The general guidelines for the construction of MLFF networks should also be followed if applicable but usually some experimentation will still be necessary Besides the time series itself other data or indicators may also be useful to be improving the accuracy of forecasting For example the use of moving averages trading volume momentum and other relevant data can sometimes improve forecasting accuracy significantly (Patterson et al 1993)

3 4 Simulation Results with BPTT

Example 1 Change in Business Inventories

The testing MSE obtained for different orders chosen is tabulated below This variation of testing MSE with respect to the order chosen is plotted graphically

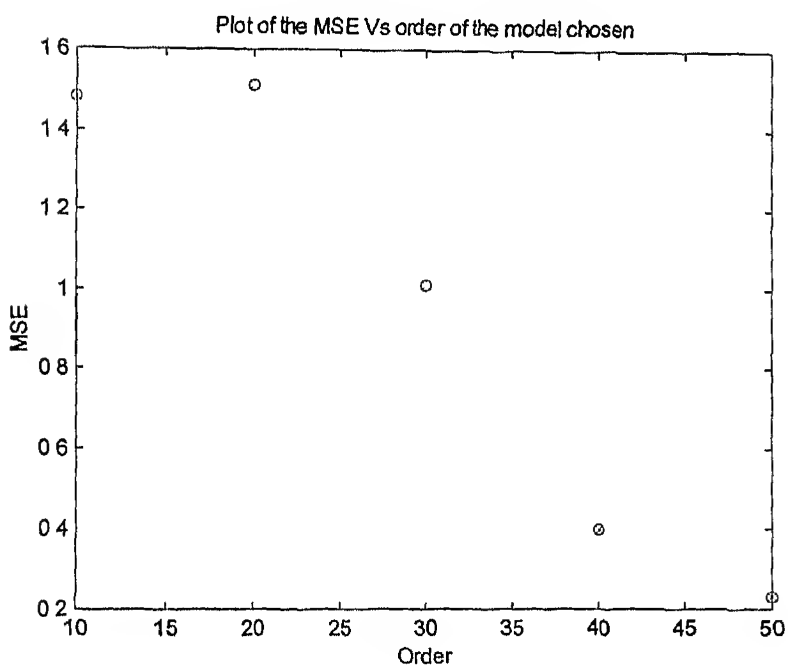
Taps	M S E
2	15 7178
5	15 8270
10	6 2372
15	7 8128
28	3 0232
35	1 1022



Example 2 Save Rate data

The testing MSE obtained for different orders chosen is tabulated below This variation of testing MSE with respect to the order chosen is plotted graphically

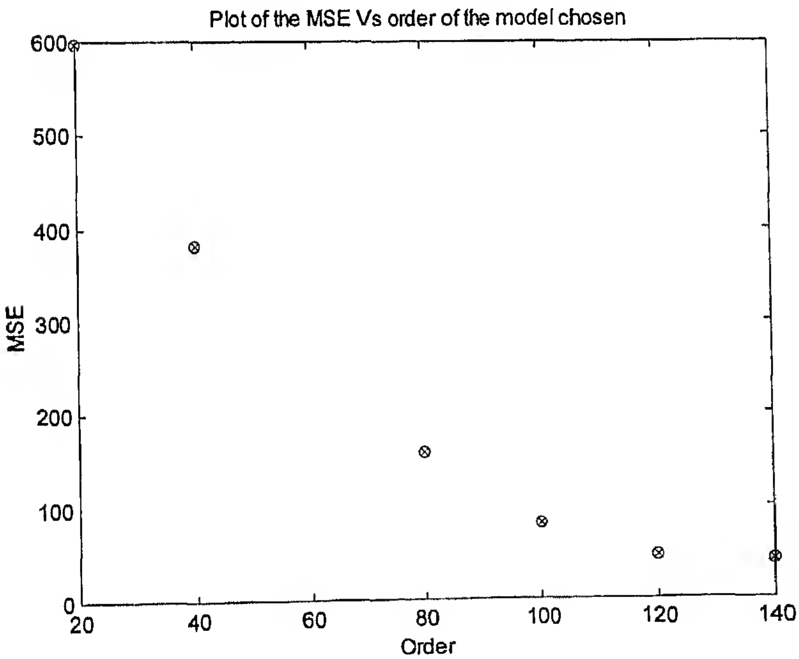
Taps	M S E
10	1 4804
20	1 5117
30	1 0077
40	0 3974
50	0 2314



Example 3 Sunspots data

The testıng MSE obtaıned for dıfferent orders chosen ıs tabulated below Thıs varıatıon of testıng MSE wıth respect to the order chosen ıs plotted graphıcally

Taps	M S E
20	97 0311
40	82 5108
80	58 3691
100	50 8420
120	46 3247
140	41 7216

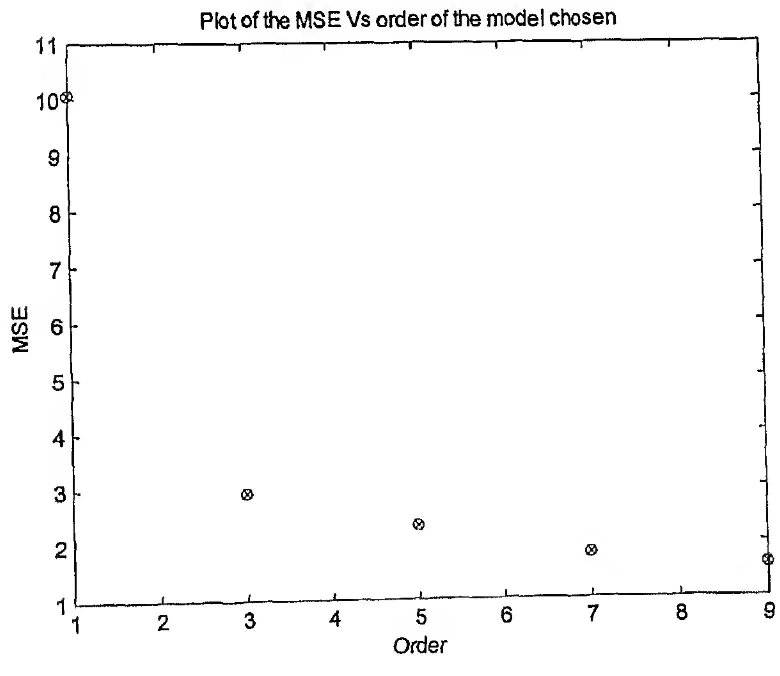


3 5 Simulation Results with TDNN

Example 1 Change in Business Inventories

The testing MSE obtained for different orders chosen is tabulated below This variation of testing MSE with respect to the order chosen is plotted graphically

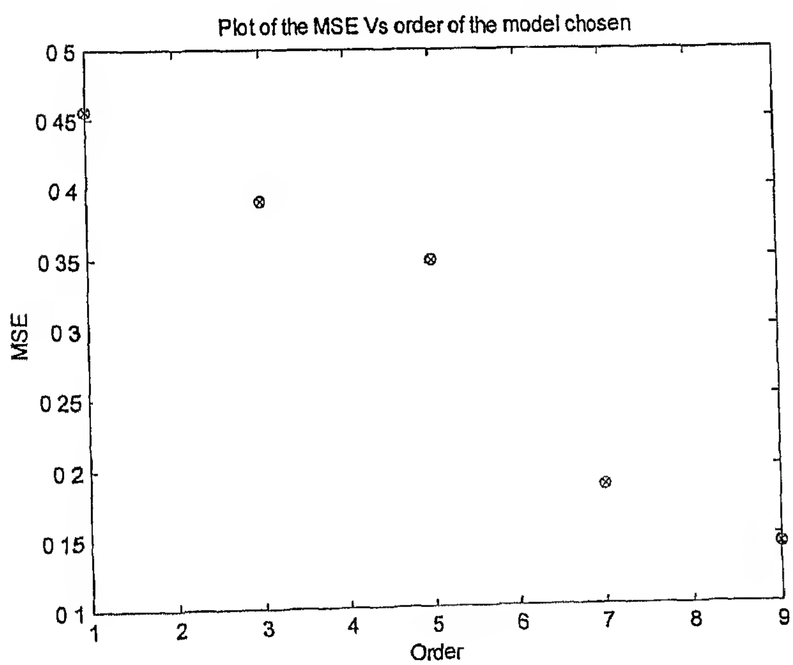
Taps	M S E
1	10 072
3	2 922
5	2 308
7	1 813
9	1 584



Example 2 Save Rate data

The testing MSE obtained for different orders chosen is tabulated below This variation of testing MSE with respect to the order chosen is plotted graphically

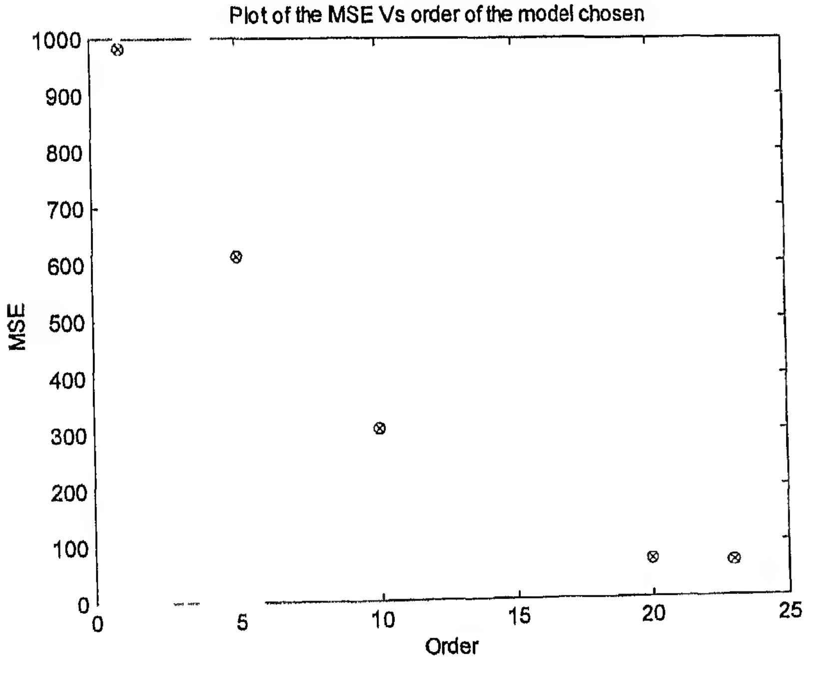
Taps	M S E
1	0 456
3	0 391
5	0 348
7	0 186
9	0 143



Example 3 Sunspots data

The testing MSE obtained for different orders chosen is tabulated below This variation of testing MSE with respect to the order chosen is plotted graphically

Taps	M S E
1	982 31
5	615 87
10	308 59
20	68 28
23	64 92



Chapter 4

Results and Discussions

4 1 Introduction

The results of the statistical and Neural Network methods are shown for three examples namely

Example 1 Change in business Inventories data

Example 2 Save rate data

Example 3 Sunspots data

The comparison of all these methods are discussed at the end of the chapter

4 2 EXAMPLE 1 Business Inventories

4 2 1 About the data

The data in this example is the change in business inventories stated at annual rates in billions of dollars The 60 observations cover the period from the first quarter of 1955 through the fourth quarter of 1969

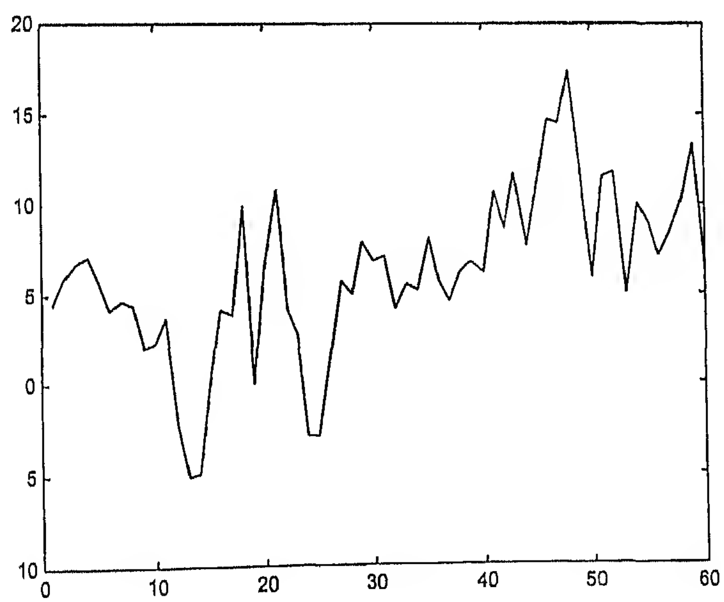
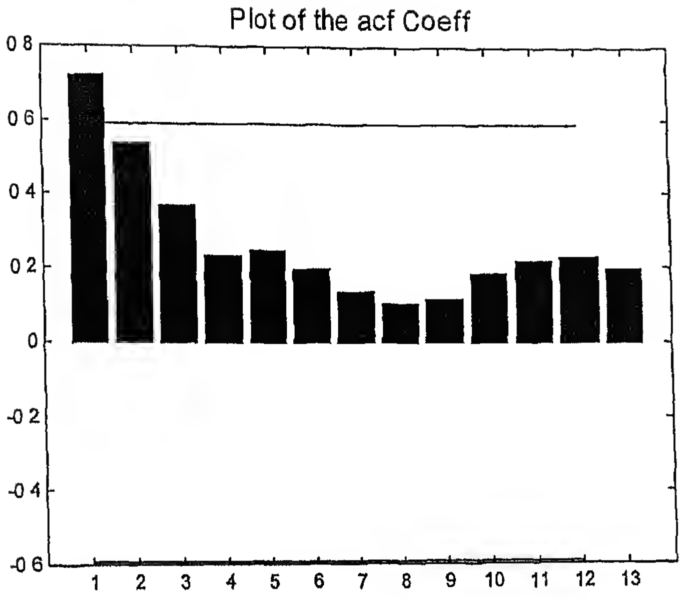
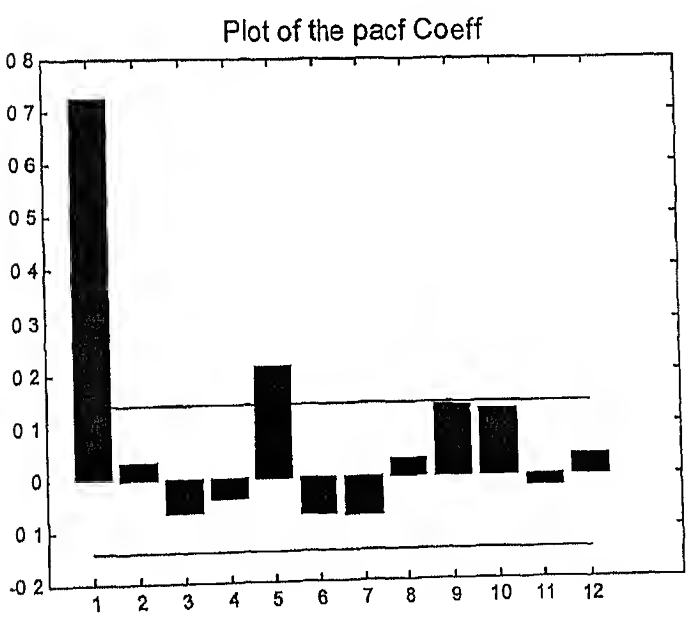


Fig 4 1 The change in the business rate data

4 2 2 UBJ Model





4 2 2 UBJ Model

Training set 0 – 50
Testing set 50 – 60
Model AR(1)

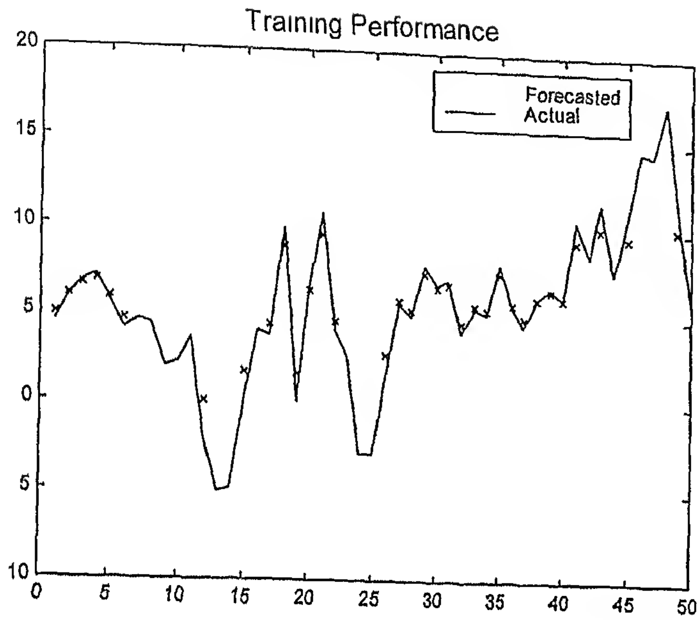


Fig 4 2 Training performance using UBJ model

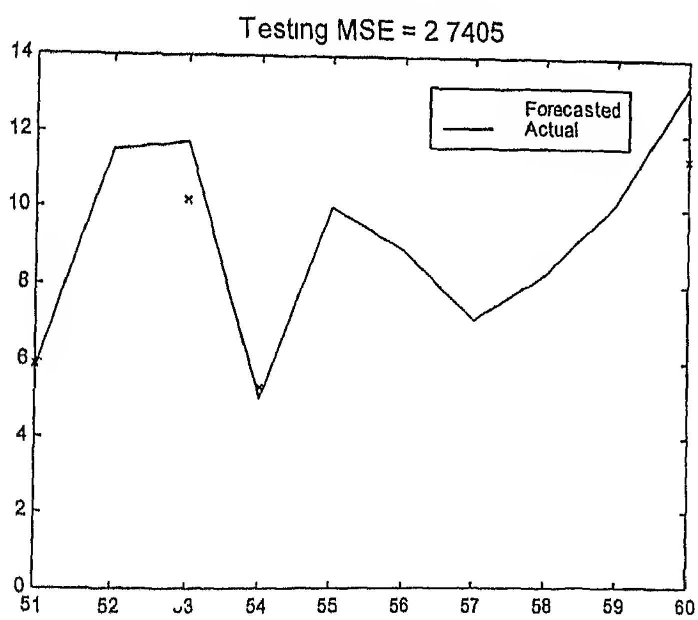


Fig 4 3 Testing performance using the UBJ model

4 2 4 Time Delay Neural Network

No of Iterations 50 000
No of hidden layers 15

Learning factor 0 000015
No of taps 7

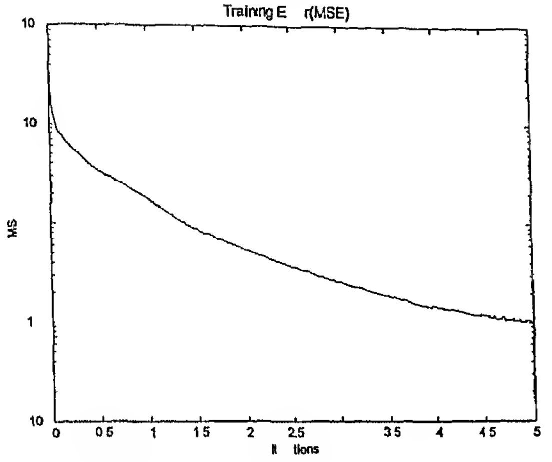


Fig 4 6 Training Error for TDNN

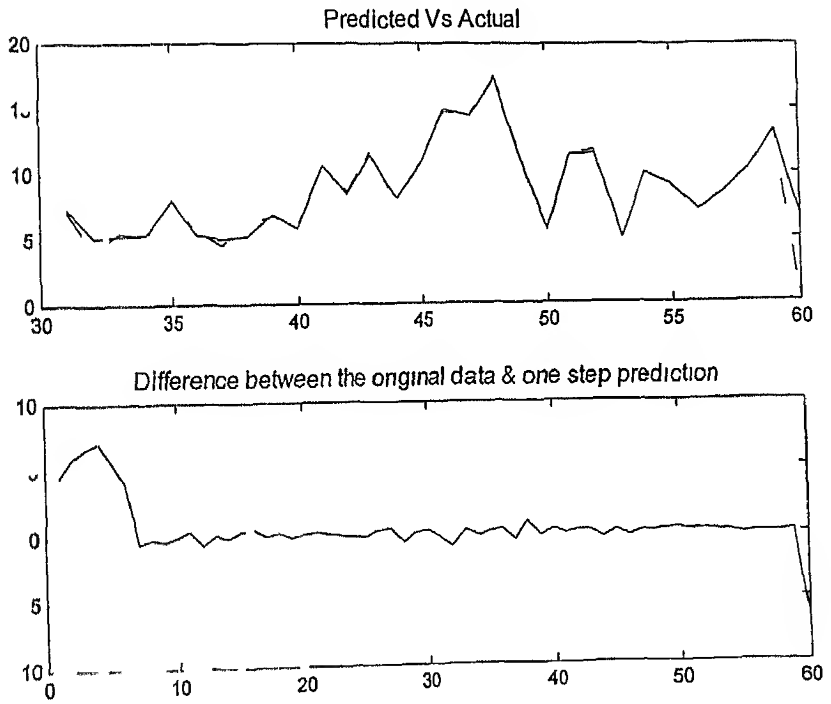


Fig 4 7 Testing Performance using TDNN

4 3 EXAMPLE 2 Save rate data

4 3 1 About the data

The saving rate is personal saving as a percent of disposable personal income Some economists believe shifts in this rate contribute to business fluctuations In this example 100 quarterly observations of the U S saving rate for the years 1955 1979 is analyzed

x axis saving rate
y axis time(quarterly)

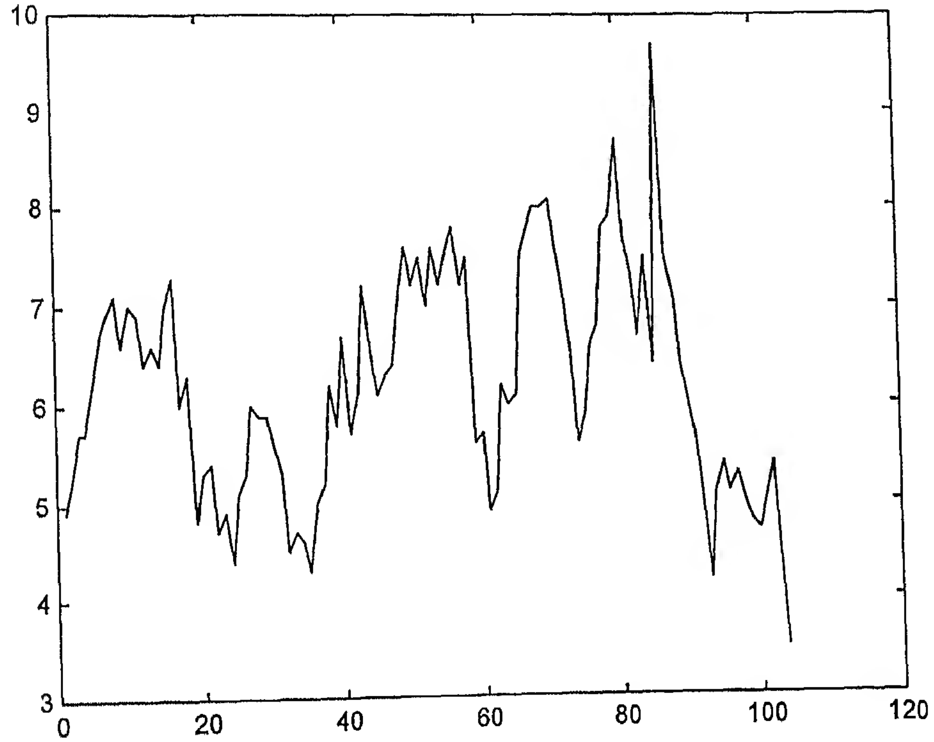
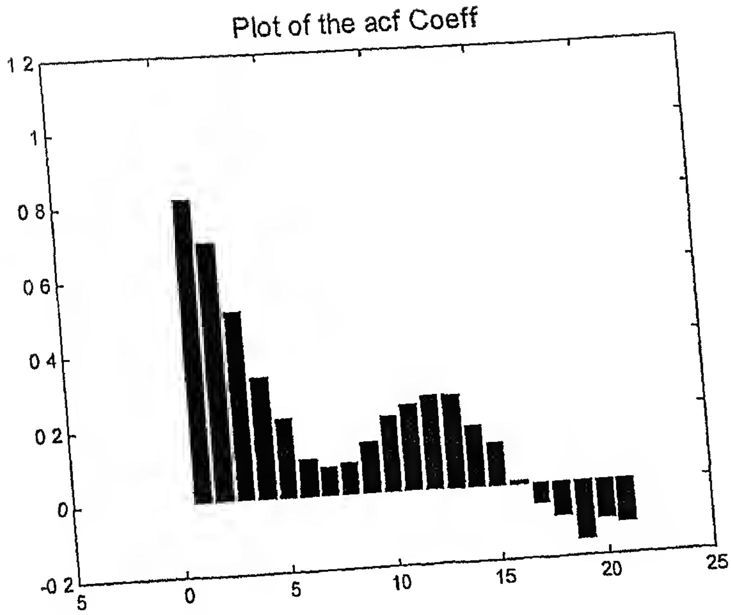
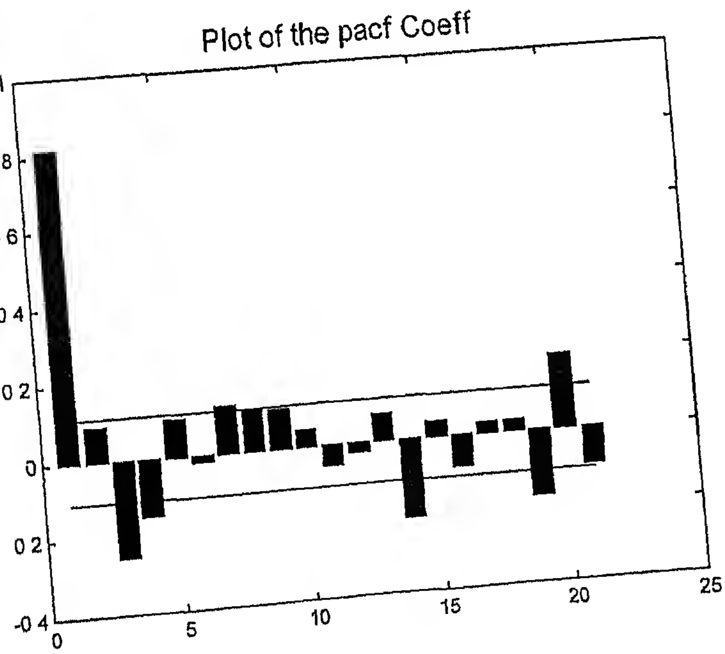


Fig 4 8 The Save rate data

UBJ Model


Plot of the acf Coeff
1 2
1
0 8
0 6
0 4
0 2
0
-0 2
5
0
5
10
15
20
25


Plot of the pacf Coeff
1
0 8
0 6
0 4
0 2
0
0 2
-0 4
0
5
10
15
20
25

4 3 2 UBJ Model

Training set	0 – 85
Testing set	86 – 104
Model	AR(2)

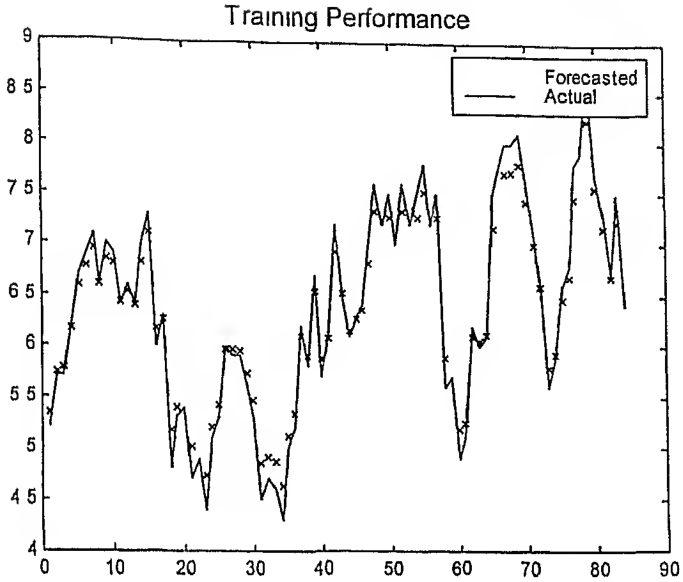


Fig 4 9 Training Performance using UBJ model

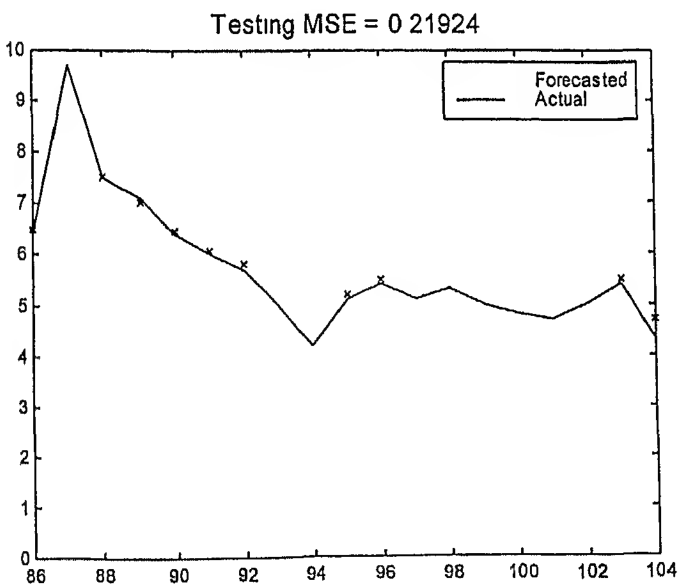


Fig 4 10 Testing Performance using UBJ model

4 3 3 Back Propagation Through Time

No of Iterations 10000
Learning rate 0 005
No of inputs taken 50

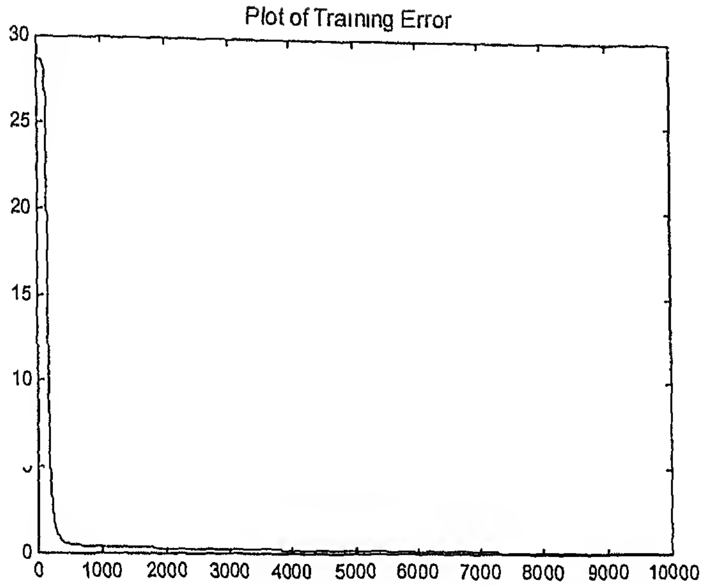


Fig 4 11 Training error (MSE) using BPTT

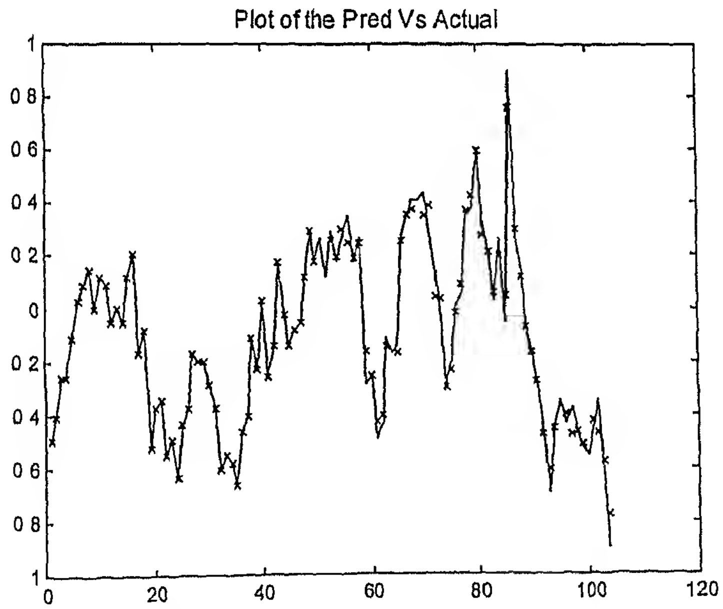


Fig 4 12 Predicted Vs Actual using BPTT

4 3 4 Time delay Neural Network

No of iterations	50000	No of hidden layers	9
Learning rate	0 0005935	No of taps	7

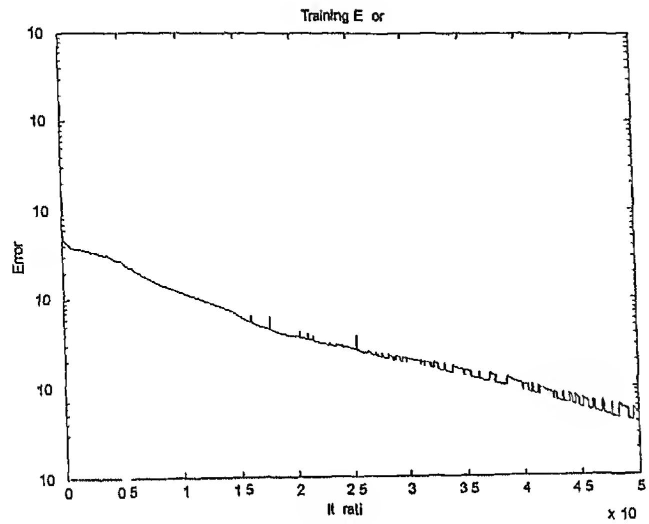


Fig 4 13 Training Error using TDNN

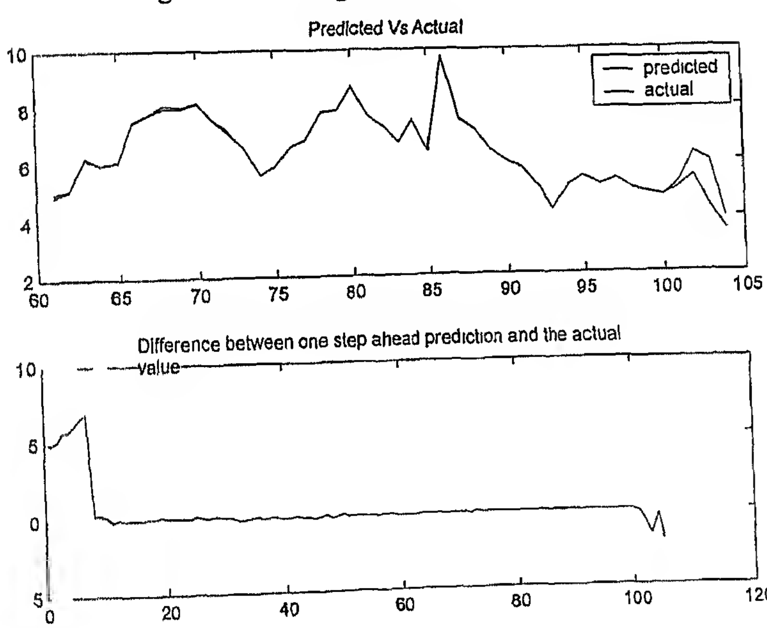


Fig 4 14 Testing Performance

4 4 EXAMPLE 3 Sunspots data

4 4 1 About the data

Sunspots are stiong magnetic field areas on the surface of the sun They are high intensity electro magnetic flaies of solar radiation of largely unknown and unpredictable causes The exact oiigin and causes are not known They have major effects on various terrestrial phenomena for example long range weather prediction telecommunications and interplanetary flight When a sunspot is in time with the earth radio signals may fade out teletype messages may be distributed Mysterious cycles are observed in sunspot data is a challenge to statisticians

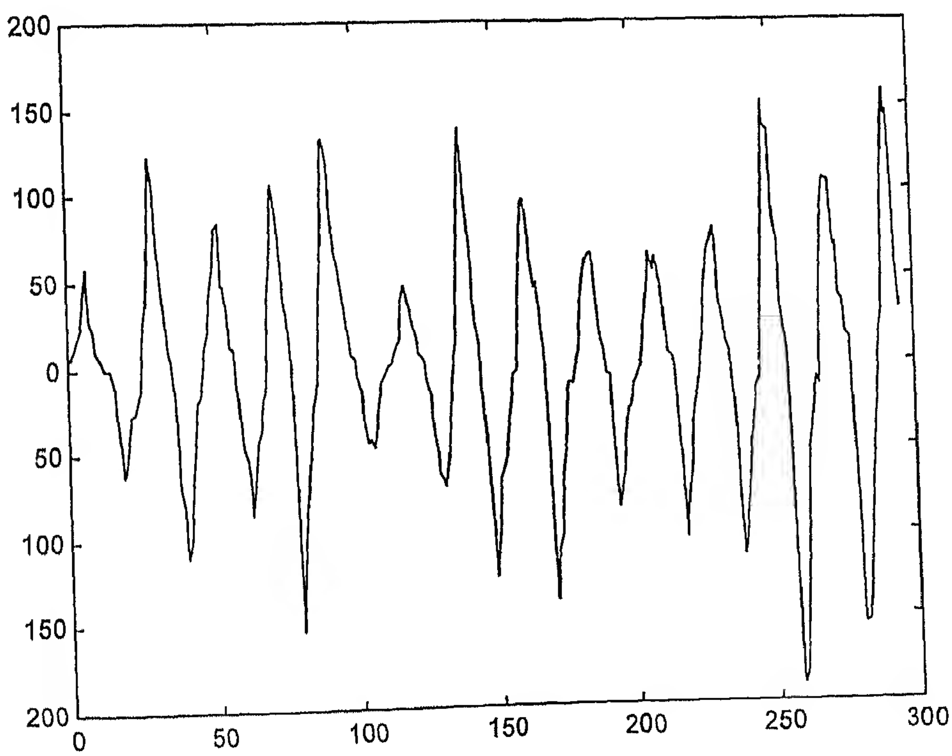
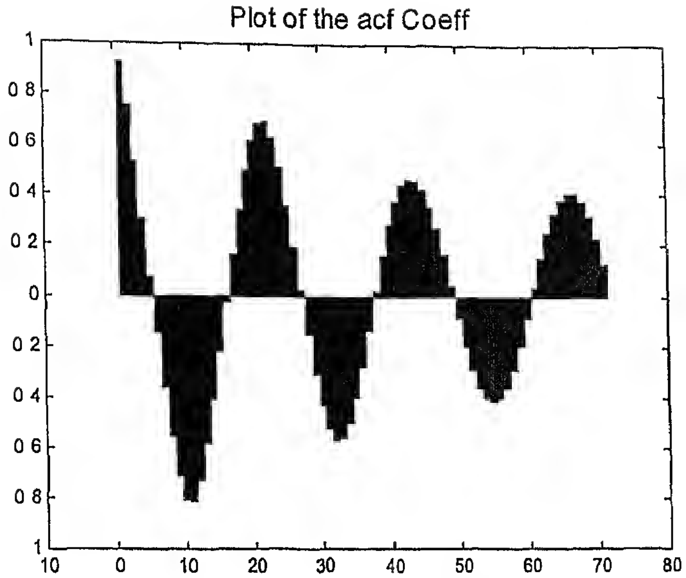
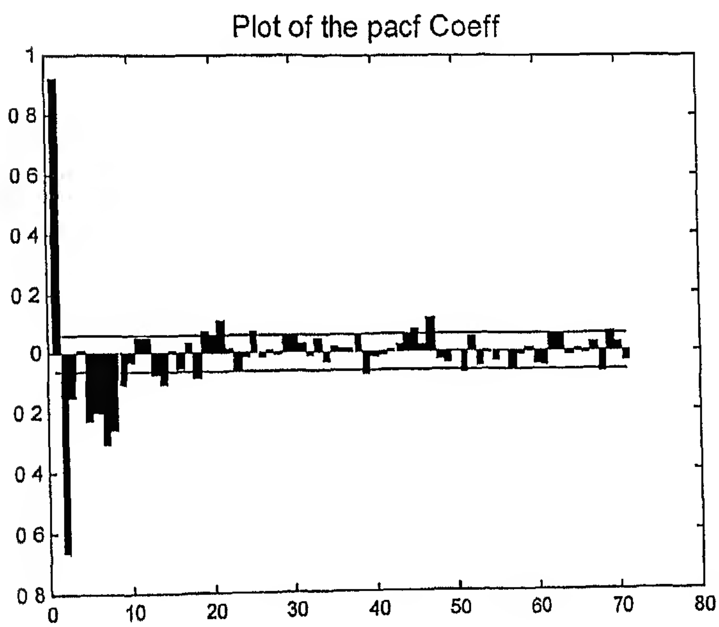


Fig 4 15 The Sunspots data

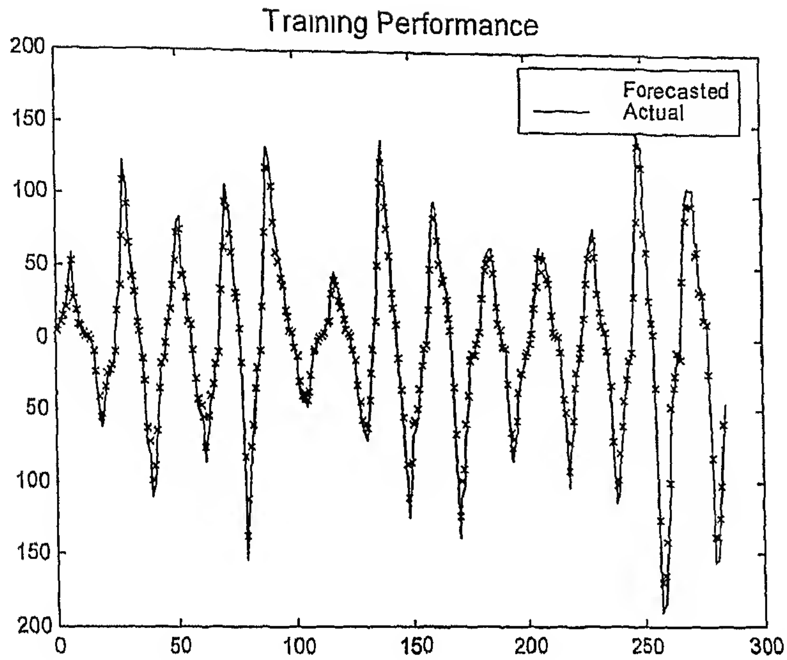
4 4 2 UBJ Model





4 4 2 UBJ Model

Traınıng set	0 285
Testıng set	286 – 295
Model	AR(1)



Γıg 4 16 The Traınıng performance usıng UBJ method

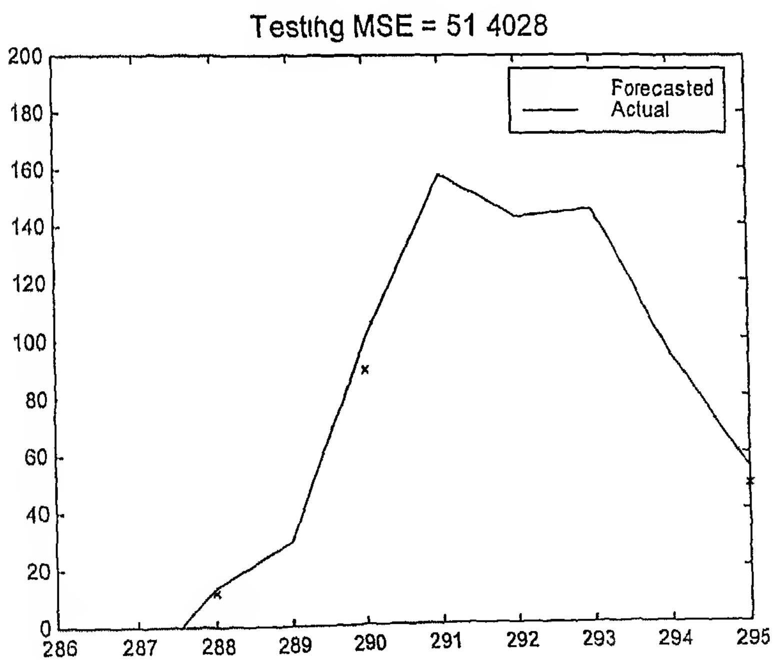
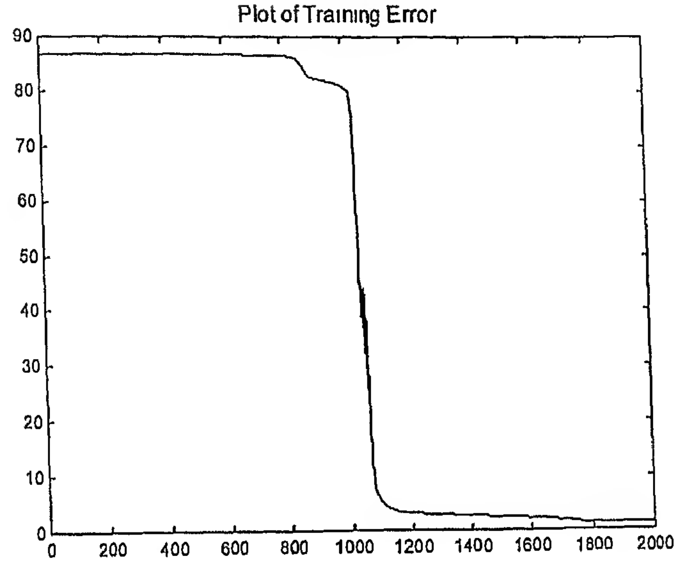


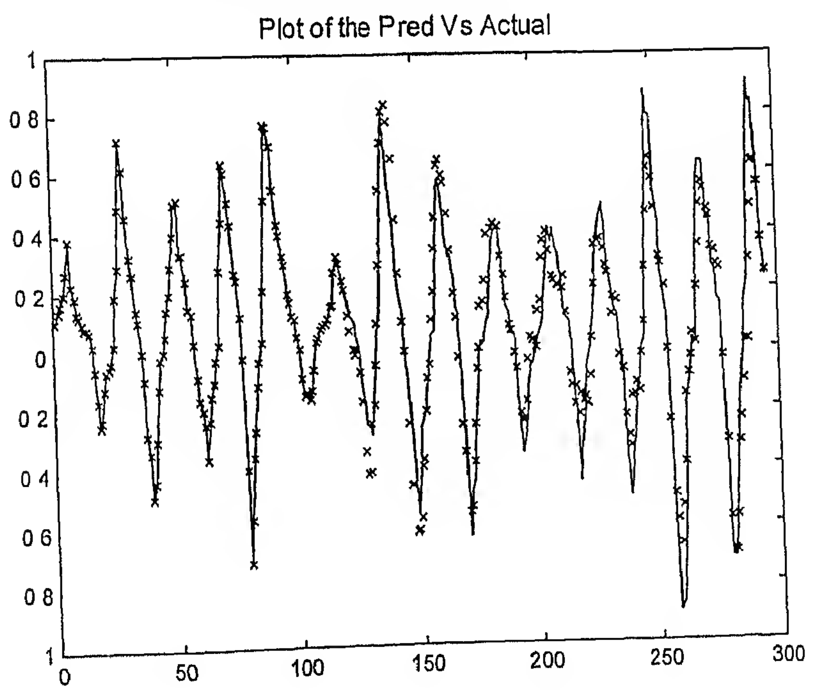
Fıg 4 17 Testıng Performance usıng UBJ Method

4 4 3 Bıck Propagatıon Thı ough Tıme

No of Iteratıons 2000
Past hıstory taken 120
Fınal error 1 2746
Learnıng factor 0 003



Гıg Tıaınıng error(MSE) usıng BPTT



Ι ıg Predıcted Vs Actual usıng BPTT

4 4 4 Time Delay Neural Network

No of iterations	60000	Learning rate	00000016
Taps	16	Hidden neurons	20

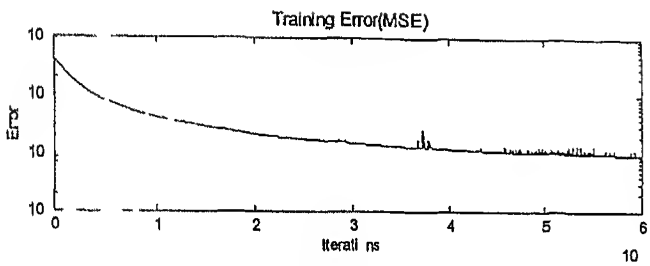


Fig 4 20 Training Error using TDNN

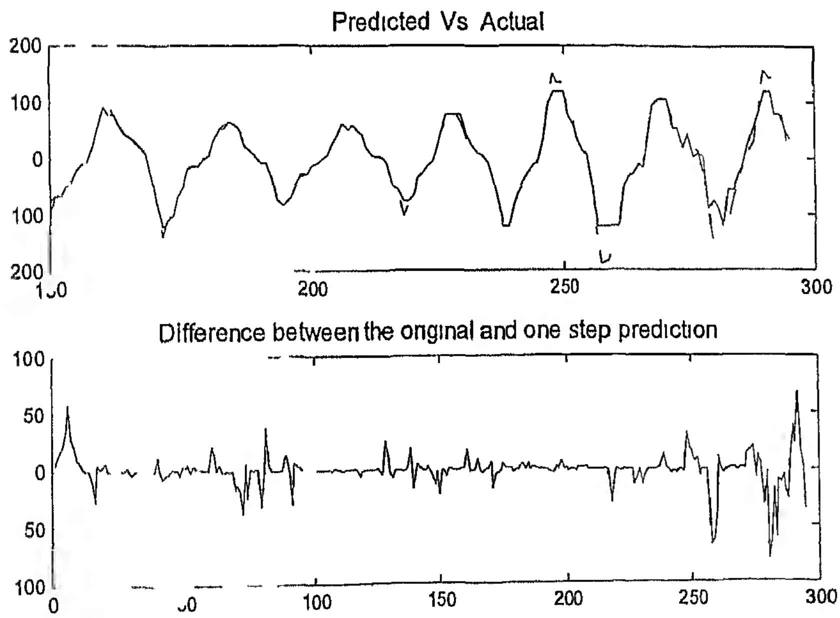


Fig 4 21 Testing Performance using TDNN

4 5 Compaı ıson of the Methods used

Γı om the sımulatıon results shown above the followıng comparısons can be made regaı dıng each method The advantages and the lımıtatıons of each method can also be done The paı ameteı s thıt aı e consıdered for the comparıson of the three forecastıng methods are

a) The accuıacy of foı ecastıng

b) The number of ıteratıons needed for traınıng

c) The amount of past hıstory needed to model ıt

Example 1

	UBJ	BPTT	TDNN
No of ıteratıons		10 000	50 000
No of past hıstorıes	1	28	7
Testıng MSE	2 7413	3 0232	1 8129

Example 2

	UBJ	BPTT	TDNN
No of ıteratıons		10 000	50 000
No of past hıstorıes	2	50	7
Testıng MSE	0 21924	0 2314	0 1859

Example 3

	UBJ	BPTT	TDNN
No of ıteratıons		2 000	60 000
No of past hıstorıes	1	120	20
Testıng MSE	51 4028	46 3247	65 2865

ARIMA models require less past history compared to the other models The accuracy of forecasting is satisfactory

BPTT models converge after large number of iterations when compared to the Time Delay Neural Networks The accuracy of prediction of the BPTT is poor when compared with that of the other two models Back Propagation Through Time needs large amounts of memory to store the intermediate weights when the data is large

TDNN models take less number of iterations when compared with the BPTT Its accuracy of prediction is better than BPTT and in comparison with that of ARIMA models

Chapter 5

CONCLUSIONS AND FUTURE SCOPE OF WORK

5 1 Conclusions

The conclusions based on the results obtained are given below

- Forecasting results are almost same for both the approaches if the time series to be modelled is not very chaotic and not very large

- Chaotic time series can be modelled more precisely by Neural Networks

- Modelling a time series using Neural Network methods is more tedious and takes a lot of time for choosing the appropriate model

5 1 Scope for Future Work Scope

- Techniques such as fuzzy clustering can be used for pre processing of the data before doing the analysis This improves the overall accuracy of forecasting

- The Neural Network model chosen can be optimally designed by employing pruning and optimisation methods

- Econometric methods can be developed using Neural Networks Multivariable analysis should be done for this purpose

References

1 Box G E P and Jenkins G M (1976) *Time Series Analysis Forecasting and Control* Holden Day San Francisco CA

2 D Rumelhart D Hinton and G Williams (1986) Learning internal representations by error propagation' *Parallel Distributed Processing Explorations in the Microstructures of Cognition* Vol 1 D E Rumelhart et al (Eds)

3 Haykin S *Neural Networks A comprehensive Foundation* IEEE Press Macmillan New York NY 1994

4 Sharda R and Patel R B (1992) Connection approach to time series prediction an empirical test *Journal of Intelligent Manufacturing* 3,317 323

5 Werbos P J (1990) Backpropagation Through Time What It Does and How to Do It *Proceedings of the IEEE*, Vol 78 No 10

6 Hassoun M H (1995) *Fundamentals of Artificial Neural Networks* PHI New Delhi

7 Zurada , J M *Introduction to Artificial Neural Systems* Jaico Publications

8 Drossu, R and Obradovic, Z Rapid Design of Neural Networks for Time Series Prediction IEEE Computational Science and Engineering

9 Wan, E A , (1994) Time series prediction by using a connectionist network with internal delay lines, In Time Series Prediction Forecasting the future and Understanding the Past, Addison Wesley

10 Patterson D W *Artificial Neural Networks Theory and Applications* Prentice Hall

11 Pankratz A *Forecasting with Univariate Box Jenkins Models Concepts and Cases* John Wiley & sons

12 Wadsworth H M *Handbook of Statistical Methods for Engineers and Scientists* 2nd edition Mc Graw Hill

Date Slip

This book is to be returned on the date last stamped